Evaluating gene flow and habitat connectivity between *Salamandra salamandra* lineages across heterogeneous landscapes

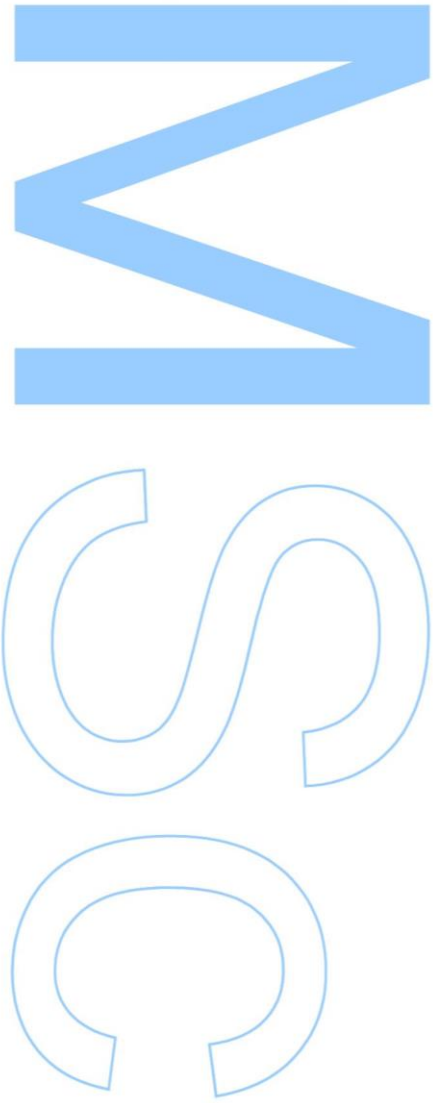
Bernardo Flores Antunes
Mestrado em Biodiversidade, Genética e Evolução
Departamento de Biologia
2016

Orientador
Guillermo Velo-Antón, Investigador, CIBIO-InBIO

Coorientador
Helena Gonçalves, Post-doc, CIBIO-InBIO
Iñigo Martínez-Solano, Investigador, MNCN

Todas as correções determinadas pelo júri, e só essas, foram efetuadas.
O Presidente do Júri,

Porto, _____/_____/_____



Agradecimentos

Ao meu orientador Guillermo Velo-Antón, pela oportunidade de realizar este trabalho, e por me orientar da melhor maneira possível. Disponibilizando todos os meios necessários para realizar esta tese (e.g. conhecimento científico, financiamento para análises de laboratório, aluno de doutoramento André Lourenço, computadores alheios, etc) e por acompanhar de perto a evolução da mesma, sempre com a maior disponibilidade.

Aos meus co-orientadores Helena Gonçalves e Iñigo Martínez-Solano, principalmente por me terem disponiblizado a maior parte das amostras utilizadas neste estudo.

Ao Guilherme Dias, que foi o responsável pelo trabalho de laboratório que resultou na extensa base de dados que me foi disponibilizada.

Ao André Lourenço por me orientar (mesmo sem o estatuto de orientador) ao longo de todo este trabalho, e em todas as suas vertentes, desde o laboratório, análises de dados, até á parte escrita.

A todo o pessoal do CTM, em particular à Patrícia Ribeiro por toda a ajuda no laboratório, e como qualquer trabalho que envolva microssatélites à Susana Lopes.

Aos Biodeserts, pela experiência de pertencer a um grupo de investigação.

Salamander Group – Guillermo Velo-Antón, André Lourenço, Kevin Mulder e Marco Dinis

Inês

Família e amigos!

Resumo

O sul da Península Ibérica é uma região geológica, climática e topograficamente heterogénea, associada com eventos de especiação e descontinuidades filogeográficas numa grande variedade de taxa. Um exemplo é a salamandra-de-fogo (*Salamandra salamandra*), que apresenta uma elevada diversidade nesta região, com cinco subespécies co-ocorrendo numa área relativamente pequena. *Salamandra s. crespoi* e *S. s. morenica* apresentam uma distribuição relativamente contínua (com ocorrências pontuais de *S. s. galaica* e *S. s. bejarae*, uma vez que este é o limite sul das suas áreas de distribuições) ao longo de um eixo oeste-este, do Alentejo em Portugal até Múrcia em Espanha, com uma zona potencial de contacto perto da fronteira luso-espanhola; enquanto *S. s. longirostris* está geograficamente isolada dos outros taxa ao sul da bacia do rio Guadalquivir.

Estudos anteriores baseados em análises filogenéticas de sequências de DNA mitocondrial (mtDNA) identificaram *S. s. longirostris* como o grupo irmão de todas as restantes linhagens de *S. salamandra* o que, juntamente com a sua distribuição alopátrica e morfologia específica, levou alguns autores a proporem *S. s. longirostris* como uma espécie distinta. Contudo, nunca foi avaliada a existência de fluxo génico nem de diferenças ambientais entre estas subespécies.

Neste trabalho, combinámos dados genéticos (mtDNA, genealogias nucleares e microssatélites) e espaciais (clima, cobertura vegetal e relevo) numa abordagem integrada com o objectivo de: (1) delinear a distribuição das subespécies estudadas; (2) avaliar a ocorrência de fluxo génico entre elas; (3) determinar a existência de potenciais diferenças ambientais entre as duas principais linhagens do sul da Península Ibérica; (4) caracterizar os padrões de diversidade e estrutura genética dentro de cada subespécie, e (5) avaliar a conectividade entre as populações de *S. s. longirostris*.

Embora as análises do DNA mitocondrial tenham recuperado cinco clados distintos correspondentes às cinco subespécies descritas, as genealogias nucleares mostraram a partilha de alelos entre todos os taxa. Este resultado poderá indicar a existência de separação incompleta de linhagens ou introgressão influenciada por mecanismos evolutivos específicos (como por exemplo, dispersão enviesada dos sexos ou selecção mitocondrial). As análises de microssatélites não detectaram nenhuma evidência para a ocorrência de fluxo génico entre a alopatricamente isolada *S. s. longirostris* e as outras subespécies. Adicionalmente, foi detectada a existência de mistura nuclear em duas zonas de contacto distintas: uma envolvendo *S. s. crespoi* e *S. s. gallaica* e, a outra, estas duas subespécies e *S. s. morenica*, apresentando mais uma

vez um possível cenário de introgressão nuclear influenciado pela dispersão enviesada de machos.

A análise da estrutura genética das populações de *S. s. longirostris*, em conjunto com as análises de conectividade de habitats, detectaram uma população potencialmente isolada no extremo oriental da distribuição desta subespécie, enquanto para as outras populações não foi identificada ausência de conectividade. A distância geográfica foi identificada como o principal motor de diferenciação genética entre as populações de *S. s. longirostris*, porém isto poderá ser apenas o reflexo do extremo isolamento da população oriental.

Os modelos de nicho ecológico (ENMs) identificaram a bacia do rio Guadalquivir como uma área altamente inadequada para a ocorrência da espécie, separando *S. s. longirostris* das restantes linhagens de *S. salamandra*. As projecções destes modelos para o passado, juntamente com as projecções para o presente, identificaram dois refúgios no sul da Peninsula Ibérica, associados a estas duas linhagens de *S. salamandra*. Em relação às comparações de nichos, o nicho de *S. s. longirostris* foi identificado como não-equivalente aos das restantes subespécies.

Os resultados obtidos neste estudo fornecem uma visão abrangente dos processos históricos e contemporâneos que condicionam a biodiversidade no sul da Península Ibérica. Este estudo suporta a existência de múltiplos refúgios isolados no sul da Península Ibérica, que permitiram a persistência de *S. salamandra* nesta área durante as oscilações climáticas do passado, corroborando um possível cenário de contracções conduzindo ao isolamento alopátrico, associado a processos de especiação, e a posteriores expansões, resultando em zonas de contacto secundário e permitindo a ocorrência de fluxo génico. Actualmente, as zonas de contacto envolvendo subespécies com distribuições parapátricas (*S. s. crespoi*, *S. s. morenica*, e *S. s. gallaica*), mostram padrões de introgressão nuclear possivelmente relacionados com a dispersão enviesada dos machos. *Salamandra salamandra longirostris* está alopatricamente isolada das outras subespécies, tendo a bacia do rio Guadalquivir sido identificada como uma barreira ao fluxo génico e provavelmente desempenhando um importante papel na especiação de *S. salamandra* no sul da Península Ibérica.

Abstract

Southern Iberia is a geologically, climatically and topographically heterogeneous region associated with speciation events and phylogeographic breaks across a variety of taxa. The Fire salamander (*Salamandra salamandra*) shows high diversity in this region, with five subspecies co-occurring in a relatively small area. *Salamandra s. crespoi* and *S. s. morenica* show a relatively continuous distribution (with punctually occurrences of S. s. gallaica and S. s. bejarae, as this is the southern limit of their ranges) along a west-east axis from the Alentejo in Portugal to Murcia in Spain, with a potential contact zone near the Portuguese-Spanish border, whereas *S. s. longirostris* is geographically isolated from the other taxa, south of the Guadalquivir river basin.

Previous studies based on phylogenetic analyses of mitochondrial DNA sequences have identified *S. s. longirostris* as the sister group to all remaining *S. salamandra* lineages, which, in addition to its allopatric distribution, and specific morphology, has led some authors to propose full species status for *S. s. longirostris*. However, no previous studies have assessed patterns of gene flow or environmental differences among neighboring *Salamandra* subspecies.

Here we combined genetic (mtDNA, nuclear genealogies and microsatellites) and spatial data (climate and landcover) in an integrative approach to: (1) delineate the distribution of the studied subspecies; (2) quantify gene flow between them; (3) assess potential environmental differences between the two main southern Iberian lineages, (4) characterize patterns of genetic diversity and structure for each southern Iberian subspecies, and (5) evaluate connectivity among *S. s. longirostris* populations.

While mtDNA analyses recovered five clades associated with the five subspecies, nuclear genealogies show allele sharing among all taxa that could be the result of incomplete lineage sorting, or introgression events, under the influence of specific evolutionary mechanisms (e.g. sex-biased dispersal or mitochondrial selection). Microsatellite analyses found no evidence of gene flow between the allopatric isolated *S. s. longirostris* and other subspecies. Additionally, nuclear admixture was detected in two different contact zones, one involving *S. s. crespoi* and *S. s. gallaica*, and another among these two subspecies together with *S. s. morenica*, presenting again a possible scenario for nuclear introgression under the hypothesis of male-biased dispersal. Genetic structure analysis among *S. s. longirostris* populations, together with habitat connectivity analysis, detected a potential isolated population in the eastern extent of *S. s. longirostris* distribution range, while for other population no lack of connectivity was identified. Geographic distance was identified as the main driver of genetic differentiation

among *S. s. longirostris* populations, however it could be a byproduct of the extremely isolated eastern population.

Ecological niche models (ENMs) for the present identified the Guadalquivir river basin as a highly unsuitable area, separating *S. s. longirostris* from neighboring *S. salamandra* lineages. Past projections of ENMs (LIG, LGM, Mid-Holocene), together with current projections, identified two isolated refugia in southern Iberia, associated with this two different *S. salamandra* lineages. Regarding niche comparisons, *S. s. longirostris* niche was identified as non-equivalent to those of other subspecies.

Our results provide a comprehensive view of the historical and contemporary processes affecting biodiversity across southern Iberian Peninsula. This study supports the existence of multiple isolated refugia in southern Iberia, allowing *S. salamandra* to persist in this area during past climatic fluctuations, fitting in a scenario of range contractions leading to allopatric isolation (associated with speciation processes), and posteriorly expansions resulting in secondary-contact zones (allowing gene flow). Currently, contact zones involving the parapatrically distributed subspecies (*S. s. crespoi*, *S. s. morenica* and *S. s. gallaica*), showing patterns of nuclear introgression possibly caused by male-biased dispersal. *Salamandra s. longirostris* was allopatrically isolated from other subspecies, with the Guadaquivir river basin identified as a barrier to gene flow, which seems to be playing an important role in species formation in southern Iberian *Salamandra*.

Table of Contents

List of Figures

Figure S3. Projections of ecological niche models, showing suitability values, for *S. s. crespoi/morenica* lineage, for the present, Mid-Holocene, Last glacial maximum (CCSM and MIROC) and Last interglacial, using 6 climatic variables (BIO3, Isothermality; BIO8, Mean Temperature of Wettest Quarter; BIO9, Mean Temperature of Driest Quarter; BIO11, Mean Temperature of Coldest Quarter; BIO12, Annual Precipitation; BIO18, Precipitation of Warmest Quarter). Standard deviation of the models represented in the small inset.

List of Tables

1. Introduction

1.1. Speciation

The study of speciation and the processes underlying it has long been subject of evolutionary studies. Speciation constitutes a mechanism responsible for the biological diversity in planet Earth, which arises from the establishment of reproductive barriers to gene flow between populations, causing genetic, ecological and phenotypic distinctiveness between them (Coyne & Orr 2004; Nosil & Feder 2012). However, an unambiguous delimitation of species is not straightforward. Several species concepts have been proposed so far (e.g. biological, ecological, evolutionary, among many others described by Mayden (1997)) but they present contrasting premises. More recently, Queiroz (2007; 2005) proposed a unified species concept to solve this problem. This unified species concept identifies species as separately evolving metapopulation lineages (which is the common element to all contemporary concepts and definitions of species), and by doing so, it deals with all the other properties that species may or may not acquire during the course of their existence (e.g. different phenotypes or ecological niches) (Fig. 1). Accordingly, proper understanding of speciation and its underlying processes is dependent upon the integration of crucial data, namely genetic and ecological information.

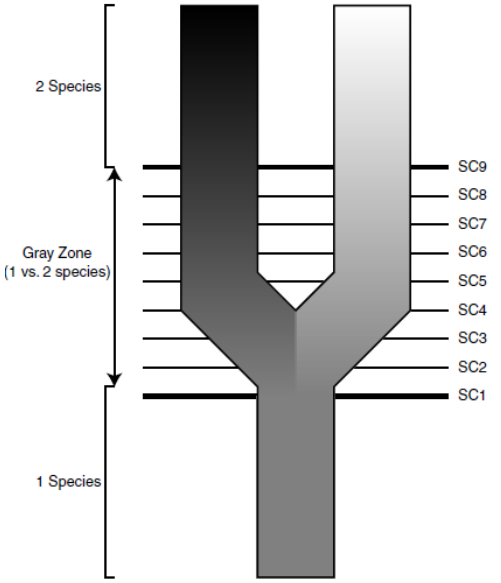


Figure 1. A highly simplified representation of the process of metapopulation lineage divergence (speciation) illustrating the conflicts caused by adopting different contingent properties of metapopulation lineages as necessary properties of species. SC1 to SC9 represent species concepts, associated with different characteristics of the species. Figure from Queiroz (2007).

From the genetic point of view, divergence can occur in response to several forces and mechanisms, such as gene flow, natural selection, sexual selection, genetic drift and chance events (e.g. mutations, chromosomal rearrangement and polyploidy; Butlin *et al.* 2012; Futuyma 2013). Gene flow (movement of genes from one population to another) is obviously a central process in speciation, as reproductive isolation is defined by the absence or restriction of gene flow between populations (Seehausen *et al.* 2014). For gene flow to occur, individuals must disperse between populations and successfully reproduce with local individuals. Consequently,

gene flow is affected by extrinsic geographical and ecological factors combined with intrinsic biological factors.

Regarding geographic settings, researchers have categorize three different modes of speciation: allopatric, sympatric, and parapatric (Fig. 2; Futuyma 2013). Allopatric speciation occurs when populations become reproductively isolated, after the appearance of a geographic barrier to gene flow (e.g. Ponce de León *et al.* 2014; Fig. 2, A). Essentially two allopatric speciation modes are recognized - speciation by vicariance (when two large populations diverge), and peripatric speciation (divergence of a small population from a larger one; Fig. 2, B). Parapatric speciation occurs when adjacent spatially distinct populations, with some gene flow, become reproductively isolated (e.g. Fisher-Reid *et al.* 2013; Fig. 2, C). Sympatric speciation is the evolution of reproductive isolation within a single, initial panmitic population (population with random mating) (e.g. Barluenga *et al.* 2006; Fig. 2, C).

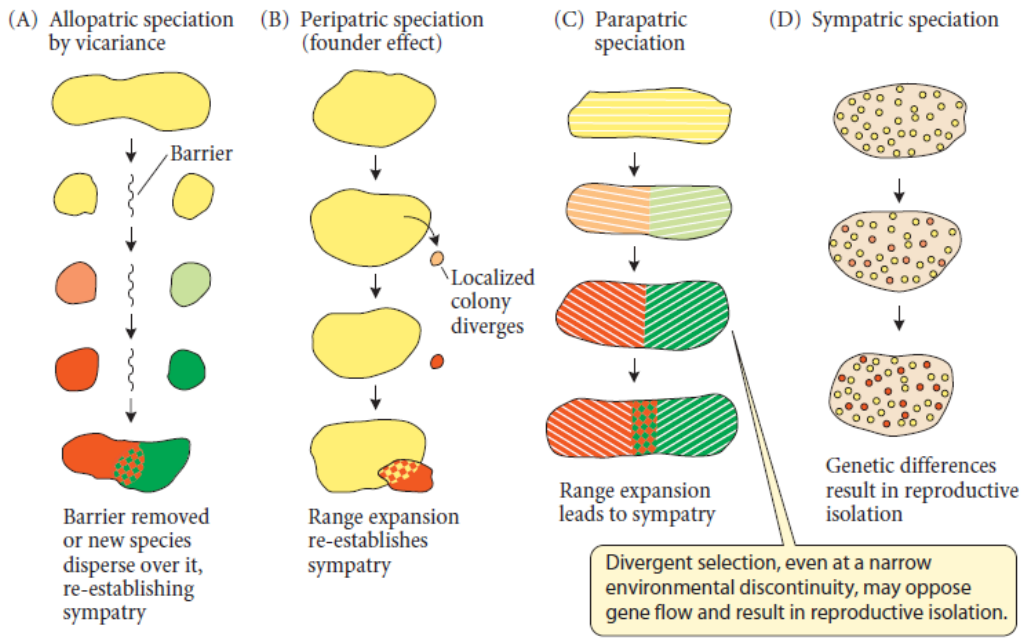


Figure 2. Schematic diagram showing the successive stages of the different modes of speciation that differ in their geographic setting. (A) Allopatric speciation by vicariance; (B) Allopatric speciation due to founder effect or Peripatric speciation; (C) Parapatric speciation; (D) Sympatric speciation. Figure original from Futuyma 2013.

Natural selection processes have been suggested to drive speciation. With habitat heterogeneity recently emphasized as a dominant force driving population divergence (Butlin *et al.* 2012). Here we will first go through ecological niche concepts, and then move to the role of ecology in speciation.

Ecological niches were first defined by Grinnell (1917), as the subdivision of habitat containing the environmental conditions that allow individuals of species to

survive and reproduce. Since then, several authors proposed new concepts, trying to improve the first definition. For instance, Elton (1927) highlighted the importance of biotic factors in ecological niches (competition, predation, parasitism, symbiosis), and Hutchinson (1957) stated that species, and not environments, have niches. The relation between some of the different concepts its represented in the BAM (biotic, abiotic, movement) diagram (Fig. 3; Soberón 2007). Considering all the concepts present in the diagram, we can think about the species ecological niche as a region containing the environmental conditions that allow individuals of species to survive and reproduce, regarding historical, geographical, and biotic factors that may be limiting its distribution (Barbosa et al. 2012).

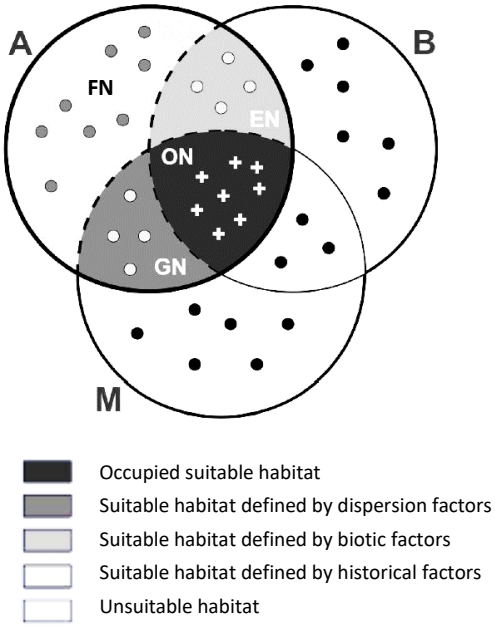


Figure 3. BAM (biotic, abiotic, movement) diagram (see Soberón and Peterson, 2005, Soberón, 2007), which represents the theoretical environmental space divided into the three main factors that limit the distribution of a species. The suitable habitat corresponds to the area common to all three factors, which represents the occupied niche (**ON**; *sensu* Pearson, 2006). The area shared by **A** and **M** represents Grinnell's niche (**GN**). The area shared by **A** and **B** is Elton's niche (**EN**). The whole **A** area is Hutchinson's fundamental niche (**FN**). A species can live in climatically favorable regions to which it has been able to disperse and from which it is not excluded by biotic interactions. Regions that fail to meet all these conditions are not suitable for the species' presence.

Ecology is considered to play an important role in speciation processes. Usually associated with niche divergence processes, for instance, with natural selection driving adaptive divergence between lineages that inhabit different environments. Although, niche conservatism (i.e. the tendency of species to retain similar ecological niches over evolutionary time scales), can work as a key factor to initially isolate populations, and create new separately evolving lineages, starting the process of speciation. (Wiens 2004) Allopatric speciation, for example, can result from niche conservatism. Moreover,allopatry can itself facilitate niche divergence, as it prevents gene flow from impeding the adaptation of populations to different local environments. (Wiens & Graham 2005).

1.2. Europe and Iberian Peninsula geological and climatic events

Current distribution of Iberian species and their evolutionary histories are deeply linked to the paleographic, climatic and topographic complexity of this region.

Iberian Peninsula had a complex paleogeographic history (Hewitt 2011). With changes during the Miocene and Pliocene periods creating a succession of geographic barriers for dispersal in numerous taxa, of which some remained, and others have been eliminated through time. For instance, during most of the Miocene, the area today occupied by the Guadalquivir river basin was occupied by a wide marine channel, known as the Betic Strait. This area has been associated to speciation processes of some taxa (Arntzen & Garcia-Paris 1995; García-París et al. 1998; Steinfartz et al. 2000), both during the Miocene (with the strait separating northern Iberia from the Betic – Riffean Massif), and since the late Miocene (after the strait was progressively closed, and the Guadalquivir river basin formed) (Fig 4.). The closing of the Betic Strait was followed by the desiccation of the Mediterranean basin, during the Messinian salinity crisis (from 5.6 – 5.3 Mya). During this period, North Africa and Europe were terrestrially connected, allowing the exchange of taxa (and gene flow) between both continents. Posteriorly, the opening of the strait of Gibraltar, and refiling of the Mediterranean Sea (ca. 5.33 Mya; Duggen et al. 2003), constituted a major barrier to dispersal of terrestrial taxa. This promoted allopatric divergence between current sister taxa inhabiting the Iberian Peninsula and North Africa (e.g. Steinfartz et al. 2000; Martínez-Solano et al. 2004; Velo-Antón et al. 2012). Nonetheless, complex patterns of single or multiple overseas colonization events between Africa and Iberia have also been suggested for some taxa (e.g. Kaliontzopoulou et al. 2011; Velo-Antón et al. 2015).

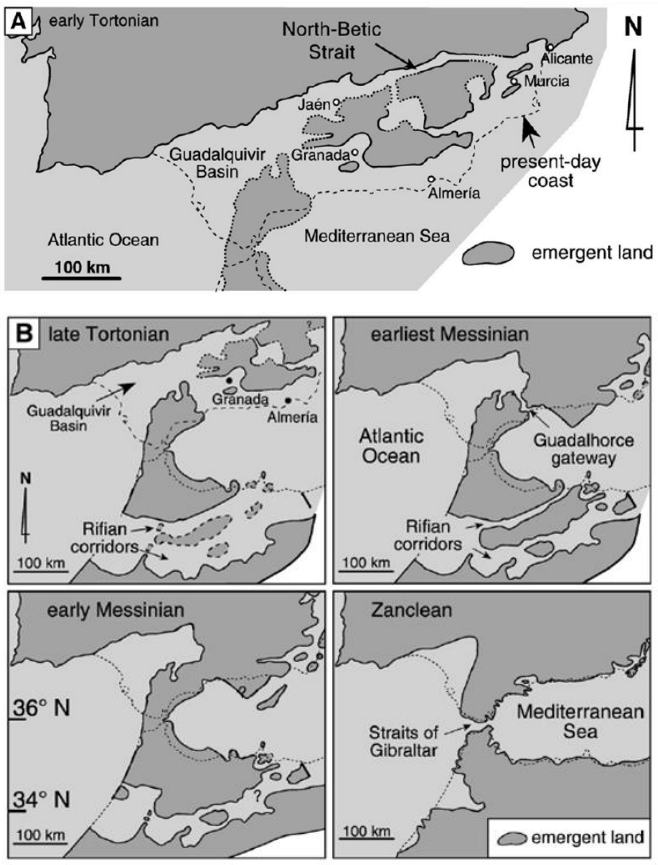


Figure 4. Southern Iberia paleogeographic reconstruction. Changes in the Atlantic-Mediterranean connections and its impact in Southern Iberia, from the Early Tortian (Late Miocene) (**A**) until the Early Pliocene **(B).** Present day shoreline identified with doted lines. Figure adapted from Martín et al. 2009.

Like past land mass changes, past climatic oscillations played a major role in shaping biodiversity patterns. In particular, the Quaternary climate was characterized by the settlement of the Arctic and Antarctic ice sheets (ca. 2.4 Mya). This period is currently acknowledged as the start of the Ice the Ages. These Ice Ages where characterized by deep climatic oscillations, with cycles of glacial periods that caused a global cooling and an increasing advances of ice sheets interchanged with warmer interglacial periods, which led to the retreat of ice sheets (Fig. 5; Hewitt 2001; 2004; 2011). During these cycles of great variability in climate and habitat, species either had to adapt to the new environmental conditions or track suitable environments by shifting their ranges, otherwise they went extinct. Therefore, species response to climatic changes would mostly depend on their life history traits. In a general view, Mediterranean (i.e. thermophilic) species would respond to interglacial or glacial periods, by respectively expanding or contracting their ranges. While, Eurosiberian (i.e. temperate) species would have the opposite response. (Abellán & Svenning 2014)

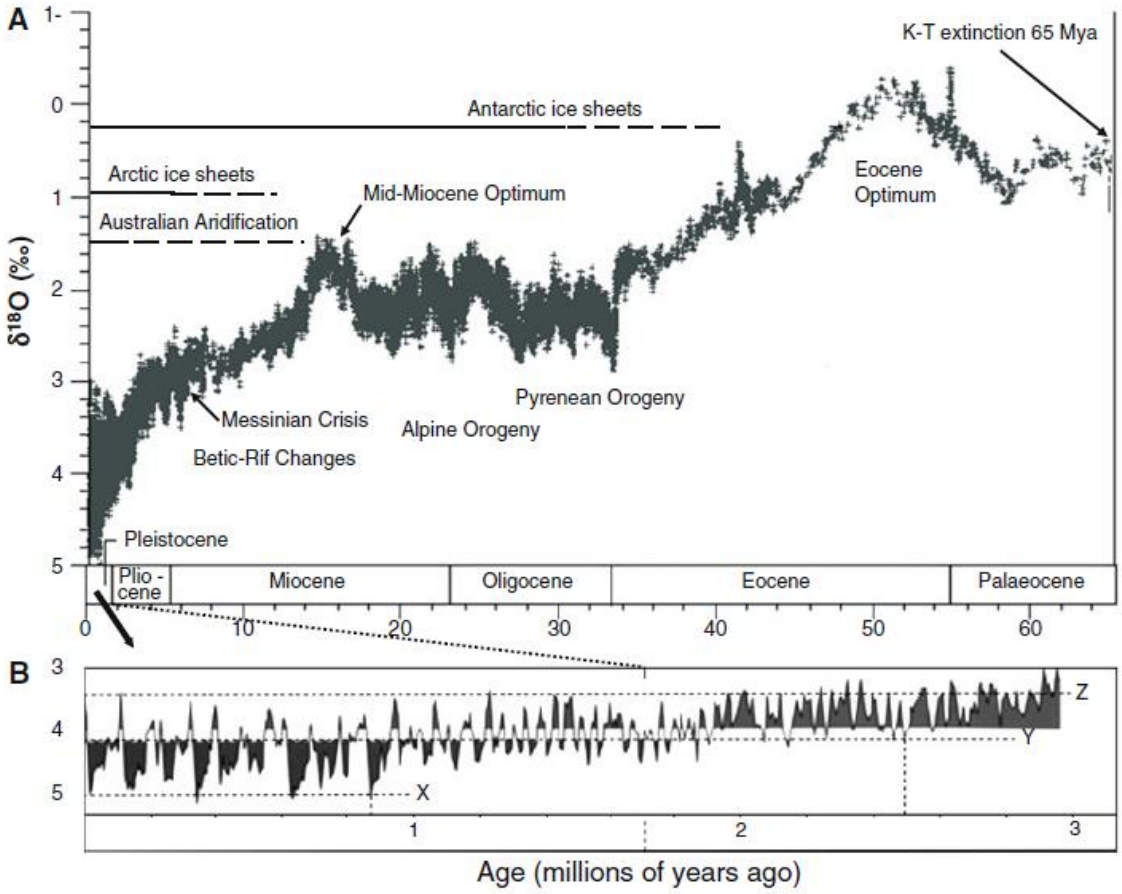


Figure 5. A time course of some geological events and climate through the Cenozoic. **A** - The climate cooled through the Cenozoic (as measured by oxygen isotope changes), with growth of Antarctic and Arctic ice sheets. Some relevant events mentioned in the text of mountain orogeny, land drift and fusion, and aridification are indicated. **B** - The climatic oscillations of the last 3 million years enlarged to show changes in major warm and cold periods; where X denotes isotope ratio near 0.9 Mya with the first 100ky ice age, Y denotes the isotope ratio some 2.5 Mya, and Z denotes the present interglacial Holocene isotope ratio. Figure obtained from Hewitt 2011.

Despite the inhospitable conditions for most species in higher latitudes, Southern European Peninsulas (Iberian, Italian and Balkan Peninsulas) constituted regions of more suitable environment, holding presumably a higher number of species. These areas are known as glacial refugia, and are largely recognized as the main Pleistocene glacial refugia for numerous Western Palaearctic species, acting both as a center of origin of endemism and as source from which species expanded with climate amelioration. In Iberia this is supported by several lines of evidence. Paleontological, palynological and paleolimnological data confirmed the persistence of numerous temperate species throughout the Ice Ages. High levels of endemism´s supported by morphological and genetic data suggest long-term survival and differentiation in this

region. Furthermore, numerous phylogeographic studies pointed Iberian Peninsula not only as the foundation for species and lineage formation, but also as a species source for the northern latitudes of Europe during inter-glacial periods. (Gomez & Lunt 2007; Hewitt 2011)

The Iberian Peninsula is a highly heterogeneous region, with several mountain ranges, major rivers, and under the influence of both north Atlantic Ocean and the Mediterranean sea, resulting in a wide range of climates, including desert Mediterranean, Alpine, and Atlantic. This wide climatic range, allows species to track their suitable climate through latitudinal and altitudinal shifts of their ranges, facilitating the existence of multiple isolated refugia within Iberia (Gomez and Lunt 2007). The hypothesis of "refugia-within-refugia" in Iberia, has been consistently suggested by several studies in this region (Centeno-Cuadros et al. 2009; Gonçalves et al. 2009; Martínez-Solano et al. 2006; Miraldo et al. 2011; Velo-Antón et al. 2012), with endemic species often exhibiting high levels of intraspecific differentiation, most likely due to their complex evolutionary history dominated by successive cycles of fragmentation, expansion and subsequent admixture of populations.

In addition, recent land use change can disrupt current biodiversity patterns in the Iberian Peninsula, with numerous species being affected by habitat change and degradation, both at local and regional scales, as it was suggested by Martins (2014), using species richness patterns and land use data in Iberia (Fig. 6).

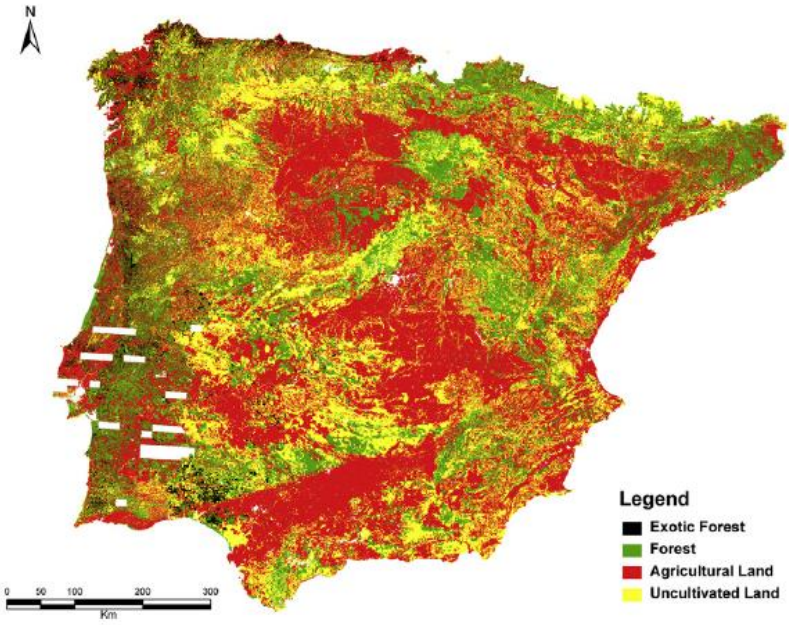


Figure 6. Land cover map of the Iberian Peninsula. Blank cells indicate missing data. Figure from Martins *et al.* 2014.

1.2.1 The study area - Southern Iberian Peninsula biodiversity

The southern part of Iberia is a geologically, climatically and topographically heterogeneous region. Its climate is mainly described as temperate and arid, greatly contrasting with the Atlantic environment observed in northern Iberia. However, the maritime influence of Atlantic Ocean and Mediterranean Sea, coupled with its complex topography, creates a wide array of areas exhibiting microclimates. These characteristics of southern Iberia suggest the existence of multiple isolated refugia during Quaternary climatic fluctuations. Current studies performed there support this hypothesis of multiple refugia, such as the two main mountain ranges of this region (Sierra Morena mountain range and Betic chain). These mountains offer a wide microclimate scope, allowing populations to survive there by altitudinal range shifts as the climate worsens or ameliorates (depending on the species ecological requirements). (Gomez & Lunt 2007; Abellán & Svenning 2014)

Amphibians and reptiles, for reasons of their low vagility, were strongly affected by the heterogeneity of southern Iberia, together with past climatic changes. This resulted in the high number of species and subspecies currently described there (Fig. 7), often exhibiting strong levels of population subdivision, as it was described in several studies (Martínez-Solano et al. 2006; Gonçalves et al. 2009; Velo-Antón et al. 2012; Santos et al. 2012). Considering this scenario, amphibians and reptiles, are recognized as good models to study biogeographic and speciation processes in southern Iberia.

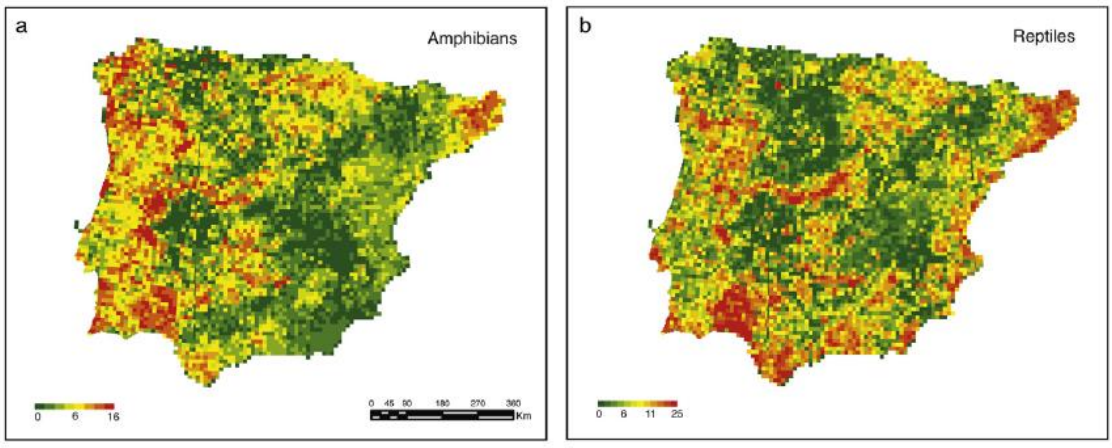


Figure 7. Map of amphibian species (a; n=31) and reptiles species (b; n=64) richness using a grain size of 10 x 10 km. Color scales are based on quantiles. Amphibians and reptile species from Pleguezuelos *et al.* 2002 and Loureiro *et al*. 2008. Figure adapted from Martins *et al.* 2014.

1.3. The Fire salamander

The Fire salamander (*Salamandra salamandra, Linnaeus*, 1758) (Fig. 8) is a terrestrial species, usually found in humid and shaded environments. They can occur over a wide range of vegetation communities, relying only on the presence of high humidity and water bodies, necessary for their common reproductive mode – ovoviviparity or larviparity – in which females give birth to many small water dependent larvae (Velo-Antón *et al.* 2007; Greven 2011).

Figure 8. *Salamandra Salamandra* pictures.

Salamandra salamandra ranges from the southwestern Iberian Peninsula to Ukraine and Greece at the easternmost extent of its range (Fig. 9). Throughout its range, coloration patterns are highly variable and it was registered intraspecific diversity of reproductive modes, namely viviparity or pueriparity (parturition of fully metamorphosed terrestrial juveniles). Also, several genetic lineages were already identified, with most of recognized subspecies (up to nine; Garcia-Paris *et al.* 2003; Beukema *et al.* 2016) being located only in the Iberian Peninsula (Fig. 9). This high intraspecific diversity and genetic differentiation is most likely related with the complex biogeographic history of Iberian Peninsula. Given the hypothesis of multiple refugia within Iberia during the climatic fluctuations of the Pliocene and Pleistocene (Gomez & Lunt 2007), it was suggested that Iberian Peninsula was the main diversification center and source of posterior colonizations across central and northern Europe (García-París et al. 2003; Velo-Antón et al. 2007). During Pleistocene climatic oscillations, *S. salamandra* populations probably survived in multiple refugia by changing their distribution range. This would involve cyclical scenarios of fragmentation of *S. salamandra* distribution as the climate worsens (potentially leading to processes of allopatric subspeciation), and expansions (resulting in contact zones) as the climate ameliorates.

Furthermore, the existence of multiple lineages and the dynamics of contact zones and introgression processes brings attention to the necessity of a reevaluation of the taxonomic validity of the several subspecies (see Pereira *et al.* 2016). Recently, by using an integrative approach which included morphological, genetic and ecological data, Beukema *et al.* (2016) revoked the subspecies status of *S. s. alfredschmidti*, which was described, mostly based on the colour patterns of restricted populations of *S. s. bernardezi*, as a valid subspecies and without acknowledging the existence of colour polymorphism in this lineage.

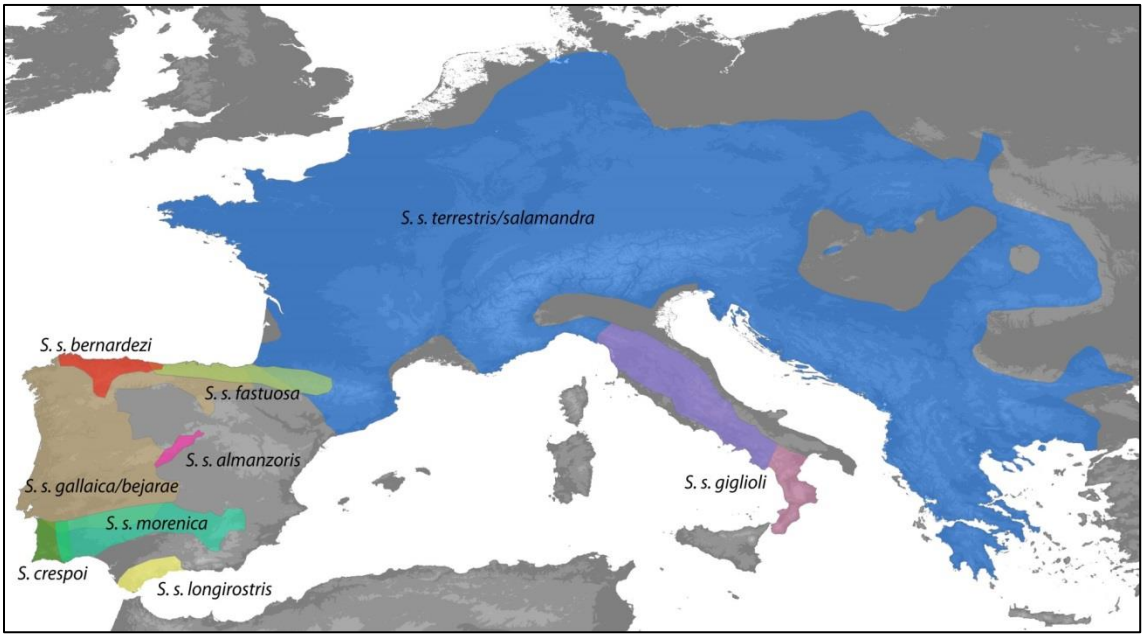


Figure 9. Distribution of *Salamandra salamandra* subspecies.

The southern third of Iberia is recognized in its majority as a Mediterranean ecogeographical region, contrasting with the majority of the regions where *S. salamandra* occurs, recognized as Atlantic regions (Pleguezuelo & Lizana 2002; Sillero *et al.* 2009). Southern Iberia is mainly occupied by three subspecies: (1) *Salamandra salamandra Crespoi;* (2) *Salamandra salamandra morenica*; and (3) *Salamandra salamandra longirostris*. The first two subspecies show a relatively continuous distribution along a west-east axis from Alentejo in Portugal to Murcia in Spain, with a potential contact zone near the Portuguese-Spanish border. On the other hand, *S. s. longirostris* is the only *Salamandra salamandra* subspecies that is geographically isolated, being distributed at south of the Guadalquivir river basin (Fig. 9). Previous studies based on phylogenetic analyses of mitochondrial DNA have identified these lineages as the oldest splits within *S. salamandra* lineages. Steinfartz *et al.* (2000) obtained a 6.3% mtDNA divergence

(control region sequences), between the *S. s. crespoi/morenica/longirostris* lineage (which were grouped in the same lineage) versus all the other *S. salamandra* lineages. Corresponding divergence times were estimated at approximately 2-4 Mya. In other study by García-París *et al.* (2003), using the mitochondrial gene (cytochrome b), *S. s. longirostris* was identified as the sister lineage to all the other *S. salamandra* lineages, with a sequence divergence of 5.1% - 5.7%. Both authors hypothesize an isolation of *S. s. longirostris* from other *Salamandra* taxa either by the Betic Strait during the Miocene, or during the Pliocene formation of the Guadalquivir river valley (2.5 – 5.3 Mya). A more recent multilocus phylogeny of *Salamandra* genus (Vences *et al*. 2014) revealed *S. s. longirostris* as the most basal mtDNA lineage in *S. salamandra*, yet analyses of several nuclear genes did not reveal genetic differentiation of *S. s. longirostris*. Another study, done by Escoriza *et al.* (2006) placed *S. s. longirostris* closer to *S. inframaculata* than to other *S. salamandra* subspecies, although they admit this may very well be an artefact.

Deep divergence at mtDNA, together with its allopatric distribution, and specific morphology (short and pointy snout, short and thick tail, and a distinct color pattern) has led some authors to propose full species status for *S. s. longirostris* (e.g. Dubois and Raffaëlli 2009). However, conflicting phylogenetic trees, and the lack of studies addressing patterns of gene flow or environmental differences with other neighboring *Salamandra* lineages led to the maintenance of its subspecies status.

1.4. Integrative analyses – combining genetic and spatial data

Considering different types of evidence obtained from independent analyses, provides the basis for testing more complex hypotheses and refining analytical models. Particularly, studies in evolutionary biology have commonly relied on morphological, molecular and ecological data (Habel *et al.* 2015).

Over the last decades, we have been able to study genetic variation directly at the DNA level, thanks to the arrival of molecular markers. Today, different types of molecular markers are available, and understanding the nature of the information provided by each of them is a crucial part of a study. Acknowledging the strengths and weaknesses of each type of molecular marker allow researchers to do an informed decision regarding the most appropriate markers to address specific questions, avoiding thus inappropriate interpretations of the results. The use of uniparentally inherited single locus markers, such as mitochondrial DNA markers, were the common method of choice for most initial phylogeographic studies. However, it is known that such data provides a very limited view of population history or can even be completely misleading (Zhang & Hewitt 2003; Sequeira et al. 2008). This can be explained by the stochastic nature of the

coalescent process, which is the major obstacle in the reconstruction of population history from a single genetic marker. Reliable studies on population structure and evolutionary history should therefore compare the phylogeographic patterns obtained from several independent loci and from different types of markers (mtDNA and nuclear DNA). Their variable mutation rates are crucial to address questions focused on different temporal scales. Indeed, mtDNA evolves approximately 5 -10 times faster than nuclear DNA (Wan *et al.* 2004; Wang 2010). This mutation rate makes mtDNA markers more powerful to infer evolutionary history in lower categorical levels such as families, genera, and species (Wan *et al.* 2004). Another widely used type of molecular marker are microsatellites - tandem repeats of 1 – 6 nucleotides found at high frequencies in nuclear DNA of several taxa. Mutations in microsatellites occur frequently by slippage and proofreading errors during DNA replication. This results in their characteristic high mutation rates (100 – 1000 faster than nDNA) that make them particularly valuable for studies of more recent processes, such as the assessment of current genetic structure (e.g. gene flow among populations, non-panmitic populations), genetic diversity and recent population history (Zhang & Hewitt 2003; Wan *et al.* 2004; Selkoe & Toonen 2006; Wang 2010).

Ecological niches models (ENM) are empirical or mathematical estimates of species ecological niches. They relate different types of eco-geographical variables (i.e. climatic, topographic or habitat variables) with the distribution of species (presence points) to identify the factors that limit and define their ecological niches and consequently, their distribution range. Lately, given their broad applicability and the increasing availability of spatial data, ENMs became a powerful tool in biodiversity studies. They allow researchers to estimate current, past and possible future potential distribution of species, characterize species environmental preferences and tolerances, and assess niche divergence/conservatism between or within species, among other applications. (Barbosa et al. 2012; Alvarado-Serrano & Knowles 2014)

Landscape genetics is a recent emergent field (Manel *et al.* 2003) in Landscape Ecology aimed at analyzing the environmental and climatic factors influencing landscape connectivity (the degree to which the landscape enhances or difficult movement), namely on gene flow patterns between individuals/populations. Specifically, it combines tools from population genetics, landscape ecology and spatial analytical techniques to explicitly quantify the effects of landscape composition, configuration, and matrix quality on microevolutionary processes, such as gene flow, drift, and selection, using neutral and adaptive genetic data (Hall & Beissinger 2014; Balkenhol *et al.* 2016; Richardson *et al.* 2016).

Landscape genetic approaches (integrating genetic and spatial data) can provide new insights into the processes influencing recent/or ongoing divergence among populations, giving us a more comprehensive view of the speciation processes. Populations connectivity/isolation can be influenced by numerous landscape components. Basically three models of isolation were defined: isolation by distance (IBD), isolation by resistance (IBR), and isolation by environment (IBE). Isolation by distance describes a pattern in which genetic differentiation increases with geographic distance, IBR incorporates landscape heterogeneity into isolation by distance, and predicts a positive relationship between genetic differentiation and the resistance distance, and IBE differentiates the strengths of habitat heterogeneity from geographic distance in genetic patterns. (McRae 2006; Wang & Bradburd 2014).

1.5. Objectives

This study assessed ecological and genetic differences between three subspecies of *S. salamandra* (*S. s. crespoi, S. s. morenica, and S. s. longirostris*) using an integrative approach that combines genetic (mtDNA, nuclear genealogies and microsatellites) and spatial data (climate and land cover). Specifically, this study aims to: (1) Delineate the distribution of the three subspecies; (2) quantify gene flow between these population groups; (3) assess potential environmental differences between them; (4) characterize patterns of genetic diversity and structure within each subspecies, and (5) evaluate connectivity among *S. s. longirostris* populations.

Overall, this study will determine the validity of each subspecies, and in particular, the potential elevation of *S. s. longirostris* as a new species by using integrative genetic and ecological analyses.

Understanding the evolutionary history and contemporary situation of these three *S. salamandra* subspecies (*S. s. crespoi, S. s. morenica, and S. s longirostris*), can give important insights into both, the complex evolutionary history of *S. salamandra*, and the historical and contemporary processes that shaped biological communities of southern Iberian Peninsula.

2. Material and methods

2.1. Study area and sampling

Sampling was conducted in areas exhibiting potential suitable habitat for the species across southern Iberian Peninsula, encompassing the entire range of the three studied subspecies (*S. s. crespoi, S. s. morenica, and S. s. longirostris*) and the southernmost ranges of *S. s. gallaica* and *S. s. bejarae*. In total, 543 fresh tissue samples (tail or toe clips) were collected mostly from larvae and some adult individuals (roadkills and living animals) in 77 locations, ranging from 1 to 23 individuals per locality (Fig. 10). Samples were georeferenced.

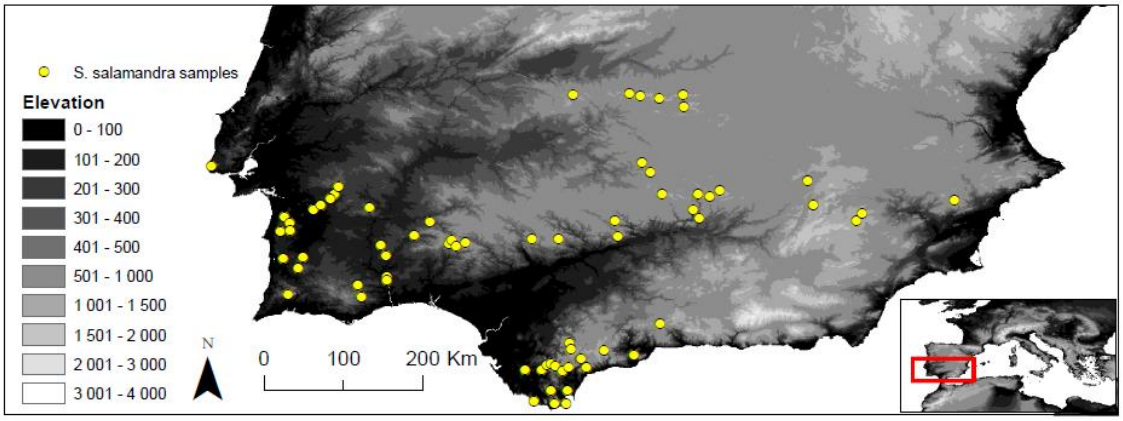


Figure 10 – Digital elevation layer of the study area (Southern Iberian Peninsula). Yellow dots represent the sampled locations (n=77), spanning in total 460 samples.

2.2. Laboratorial procedures

Total genomic DNA was extracted from 95 new samples of *S. salamandra*. These samples were combined with an existing database containing 440 fire salamander samples collected in the study area, which were previously genotyped for the same nine loci, as well as sequenced for one mitochondrial marker in a small subset. This increment in sample size was required to perform more robust downstream statistical analyses. The laboratorial procedures described below concern the former database, although protocols were identical among both databases.

DNA was extracted from tissue samples using Genomic DNA Tissue Kit (Easy-Spin), following the manufacturer's protocol. Quantity and quality of DNA extracted products were assessed by electrophoresis in a 0.8% agarose gel in a UV

transilluminator device (Bio-Rad). Successfully extracted DNA was used as template in a polymerase chain reaction (PCR) to amplify: I) one mitochondrial fragment, cytochrome b and adjacent tRNAs (*cyt-b*; ca. 1400 bp); II) ca. 700 bp of the intron of the nuclear gene β-fibrinogen (*β-fibint7*); and III) nine microsatellite loci (Sal29, Sal23, SalE7, SalE5, SalE2, SalE06, Sal3, SalE8, SalE12).

The *cyt-b* fragment was amplified for 131 samples (43 *S. s. longirostris*, 29 *S. s. morenica*, 18 *S. s. crespoi*, 29 *S. s. bejarae*, and 12 *S. s. gallaica*), using primers Glu14100L (forward, 5' GAA AAA CCA AYG TTG TAT TCA ACT ATA A 3') and Pro15500H (reverse, 5' AGA ATT YTG GCT TTG GGT GCCA 3') (Zhang *et al.* 2008), while the *β-fibint7* gene was amplified for 60 samples (18 *S. s. longirostris*, 10 *S. s. crespoi*, 19 *S. s. morenica*, and 13 *S. s. gallaica-bejarae*), using BFIB_F (forward, 5' TGG GAC TGG CAG TTG TTT AG 3') and BFIB_R (reverse, 5' TGA TTC ACG AGT TTG TTG CTC 3') (Pereira *et al.* 2016). Each PCR contained 5 µl of MyTaqTM HS Mix 2X (Bioline), 3 µl of distilled H_2O, 0.5 µl of each primer at a concentration of 10 µM and 1–2 µl of DNA extract (~50 ng/ µl), making a total volume of 10–11 µl. A negative control was used to identify possible contaminations. *For cyt-b* gene, cycling conditions were as follows: initial denaturation at 94°C for 5 min, followed by 40 cycles of 40 s at 94°C, 40 s of annealing at 51°C, elongation at 72°C for 2 min 30 s, ending with a final extension of 5 min at 72°C. PCR conditions for *β-fibint7* gene were as follows: initial denaturation at 94°C for 5 min, followed by 40 cycles of 30 s at 94°C, 30 s of primer annealing at 59°C, elongation at 72°C for 45 s, finishing with a final extension of 5 min at 72°C. Quality and quantity of PCR products were assessed by visual inspection in a 2% agarose gel in a UV transilluminator device (Bio-Rad). Also, successful amplification of the fragments of interest was determined by employing a 100-1000 DNA ladder (NZYTech) in a gel electrophoresis to allow product size comparison, thus confirming if PCR yielded the expected fragment size. Sequencing of PCR products was outsourced to Macrogen Inc. (Amsterdam, Netherlands) and Beckman Coulter Inc. (Grenoble, France). Sequencing of *cyt-b* was performed using Glu14100L and an internal forward primer (available upon request). All the obtained chromatograms were verified, aligned and corrected by eye using GENEIOUS PRO v4.8.5. A region of ca. 300 bp was trimmed in all *cyt-b* sequences after the *cyt-b* stop codon, which contained tRNA Thr, part of tRNA Pro, and a large non-coding region. This region could not be unambiguously aligned and was excluded from the analyses, resulting in a final alignment of ca. 1100 bp for the *cyt-b*.

The microsatellites employed in this study were previously developed by Steinfartz *et al.* (2004). The nine microsatellite loci were split into four distinct PCR multiplex reactions (see Table S1). These multiplexes were previously optimized and used successfully in a study conducted on fire salamanders (Álvarez *et al*. 2015),

although here only a subset of the microsatellites employed in the cited study were used. Each PCR multiplex mix contained a total volume of 10-11 µl: 5 µl of Multiplex PCR Kit Master Mix (QIAGEN), 3 µl of distilled H_2O, 1 µl of primer multiplex mix, and 1-2 µl of DNA extract (~50 ng/µl). Forward primers were labelled with fluorescent tags (6-FAM, VIC, NED or PET) for visualization purposes and reverse primers modified with a ''PIG-tail'' (GTTT) at the 5´ end to promote adenylation of the 3' end forward strand and thus, improve genotyping efficiency (Brownstein *et al.* 1996). A negative control was always used to identify possible contaminations. PCR touchdown cycling conditions were equal in three out of four multiplex reactions (panel S2, panel S4, and panel S5). The reaction started with initial denaturation at 95°C for 15 min, 19 cycles with 95°C for 30 s, 90 s of annealing at 65°C (decreasing 0.5°C each cycle), 72°C for 40 s, followed by 25 cycles of 95°C for 30 s, 56°C for 60 s, 72°C for 40 s, and ended with a final extension of 30 min at 60°C. For the panel S3, a similar protocol was performed, but with 27 cycles instead of 19, leading to a decrease in the annealing temperature to a minimum of 52°C instead of 56°C. PCR amplification quality was assessed by visual inspection in 2 % agarose gels in a UV transilluminator device (Bio-Rad). PCR products were run on an ABI3130XL capillary sequencer (Applied Biosystems) using a size standard LIZ 725 (Nimagen). Allele scoring was performed using GeneMapper version 4.0 (Applied Biosystems).

2.3. Phylogenetic analyses

In order to delimit the geographic range of each study subspecies, phylogenetic relationships were assessed for the mitochondrial (*cyt-b*) dataset, using the Bayesian approach implemented in BEAST v 1.7.5 (Drummond *et al.* 2012). JMODELTEST v.2.1.4 (Darriba *et al.* 2012) was used to test for the best fitting model of nucleotide substitution, under the Bayesian information criteria correction (BIC; HKY+G). A lognormal relaxed clock and a coalescence constant size model were used as tree priors. Markov chain Monte Carlo (MCMC) analyses were run in three independent analyses using 100 million generations, with a sampling frequency of 1000 generations and discarding 25% trees as burn-in. Parameter convergence was verified by examining the effective sample sizes (ESSs) using TRACER v1.7. The remaining trees were used to obtain the subsequent maximum clade credibility summary tree with posterior probabilities for each node using TREEANNOTATOR v1.7.5 (BEAST package). Phylogenetic relationships at nuclear *β-fibint7* were analyzed using a haplotype network. Heterozygous sequences within the nuclear *β-fibint7* fragment were phased using the PHASE algorithm as implemented in DNASP 5 (Librado & Rozas 2009). Phase probabilities parameter was set at 0.7, and all other settings were set by default. TCS v.1.21 (Clement *et al.* 2000) was used to

construct the haplotype network and applying default settings (probability of parsimony cut-off: 95%).

2.4. Microsatellites quality assessment and dataset organization

To obtain accurate results from analyses relying on microsatellite data, it is crucial to assess if loci do not violate critical assumptions of genetic models and if the sampling design is deemed adequate for a particular genetic analysis (Selkoe *et al.* 2006; Waples *et al.* 2015). These quality control procedures were performed following the order in which they are described below.

First, the total dataset was filtered based on the percentage of missing data. Specifically, samples presenting more than 40% of missing data were discarded from downstream genetic analyses. Additionally, it was demonstrated that the inclusion of related individuals may bias population and landscape genetic analyses (Goldberg & Waits 2010; Peterman *et al.* 2016). This issue is particularly relevant when sampling amphibian larvae since related larvae (i.e. full-siblings) are often spatially clustered. To reduce bias, the dataset was screened for the presence of familiar relationships (parent-offspring, full-siblings or half-siblings), using software COLONY 2.0.6.1 (Jones & Wang 2010). Runs were performed using the full-likelihood method, with high likelihood precision and medium run length, under the scenario of polygamy for both sexes. Sibship scaling option was deactivated, and no a priori information regarding known parents was provided. Other parameters were kept default. If a pair of related individuals (full-siblings or parent-offspring) had a posterior probability higher than 0.50, then only one individual from this pair would be removed. If the new dataset without highly related individuals provided results in genetic diversity and pairwise genetic differentiation analyses poorly correlated with the ones obtained by the original database (Pearson correlation coefficient, $r < 0.9$), then we would conduct further analyses using this new dataset.

Moreover, for most population-based analyses (namely genetic diversity, pairwise genetic differentiation, and landscape genetics analyses), given the scattered sampling in some areas, demes or populations (local random mating units) were defined by pooling individuals with intersecting buffers of 2 Km radius (the estimated home range area of *Salamandra salamandra*; Schulte *et al.* 2007). Additionally, only groups containing 10 or more individuals were considered demes for population genetic analyses. Scattered samples that were not assigned to any deme (i.e. individuals sampled to cover gaps between populations) were added to the subsets (see below) only for STRUCTURE analyses in order to increase sample size (see section 2.6). We

assumed HWE in the populations where these samples were collected, as the number of samples per location was not sufficient for population analyses.

In the end, the sample dataset was reorganized and/or partitioned in five distinct subsets containing: (i) the total number of demes considered in this study following the criteria enumerated in the previous paragraph (subset 1); (ii) the total number of studied demes without highly related individuals identified with COLONY software (subset 2); (iii) demes from three mtDNA lineages (*S. s. crespoi, S. s. morenica, and S. s. longirostris*) occupying the full extent of the study area (subset 3); (iv) all populations excluding demes belonging *to S. s. longirost*ris (subset 4); and (v) all the demes belonging to *S. s. longirostris* mtDNA lineage (subset 5). This was done to obtain more reliable results when addressing specific questions.

For each subset, quality control tests on microsatellites were carried out. The presence of null alleles was assessed using a maximum likelihood estimator for an individual inbreeding model, as implemented in INEst 2.0 (Chybicki & Burczyk 2009). A total of 200 000 cycles with a burn-in of 20000 cycles and each iteration saved every 50 cycles were set. MICROCHECKER v.2.2.3 (Oosterhout *et al.* 2004) was used to detect scoring errors due to stuttering or large allele dropout, with a 95% confidence level. GENEPOP 4.5.1 (Rousset 2008) was used to test significant deviations from Hardy-Weinberg equilibrium (HWE) and the presence of Linkage Disequilibrium (LD) among pairs of loci, using a Markov chain Monte Carlo (MCMC) method with 1000 batches, 10 000 iterations and 10 000 dememorization steps. The p-value of the HWE and LD multiple exact tests was corrected using the False Discovery Rate (FDR) approach (Benjamini & Hochberg 1995).

2.5. Genetic diversity

Indices of population genetic diversity were calculated for both subset 1 and subset 2 in GenAlEx 6.502 (Peakall & Smouse 2006), namely observed heterozygosity (H_O), unbiased expected heterozygosity (H_E), mean number of alleles (N_A) and mean population relatedness (R; Queller & Goodnight 1989). Estimates of unbiased allelic richness (AR) and private allelic richness (P-AR) were obtained in HP-RARE (Kalinowski 2005).

2.6. Genetic differentiation

Two measures of genetic differentiation among populations were calculated in R package *diversity* (Keenan *et al.* 2013), using 1000 bootstrap replicates, for subsets 1 and 2: (1) the pairwise F_{ST} (Weir & Cockerham, 1984); and (2) the pairwise Jost's D_{EST} (Jost 2008). Pairwise F_{ST} is known to give a more accurate estimation of population structure when both sample size and number of loci are relatively small (Gaggiotti *et al.* 1999). Its common use in population genetic studies also allows for comparisons between studies on fire salamanders using microsatellite markers. Jost's D_{EST} measures the fraction of allelic variation among populations.

Genetic structure in the present study was assessed using the Bayesian algorithm implemented in STRUCTURE 2.3.4 (Pritchard *et al.* 2000). An initial analysis was conducted in the subset 3 to investigate if *S. s. longirostris* is genetically isolated from *S. s. morenica/crespoi*. Previous genetic studies (García-París *et al.* 1998; Vences *et al.* 2014) suggested that *S. s. longirostris* constitutes a very distinct mtDNA lineage (even a different species, although this premise was not yet properly confirmed) from the remaining lineages included in this study, being the only fire salamander lineage geographically disjunct. Thus, absence of gene flow between *S. s. longirostris* and other lineages is very likely. If this scenario was confirmed, then a hierarchical STRUCTURE analysis would be conducted independently for populations from the other mtDNA lineages (subdataset 4) and for *S. s. longirostris* populations (subdataset 5) to increase the analytical power within groups.

For each STRUCTURE analysis, ten independent runs for a number of clusters (K) ranging from 1 to K (number of demes existent in each subset) were set (see section 3.2 in Results to check the number of demes in each subset). Run parameters were as follows: a burn-in period of 100000, run length of 1000000 iterations, with a correlated allele frequencies and the admixture model was set. No prior information regarding population origin was used. The most likely value of K was selected with the ΔK approach (Evanno *et al.* 2005), method implemented in the online platform STRUCTURE HARVESTER (Dent A. & VonHoldt 2012). For the best K, runs were merged and graphically displayed using Pophelper (Francis 2016).

To corroborate STRUCTURE's output regarding if *S. s. longirostris* is indeed genetically isolated, the subset 3 was reanalyzed in NewHybrids 1.1 (Anderson & Thompson 2002) to identify possible hybrid individuals between demes, using a burn-in of 1000000 sweeps followed by 10000000 sweeps. To reduce the influence of rare alleles, a uniform prior was chosen to reduce bias in the analyses.

2.7. Ecological niche-based modelling

2.7.1. Species occurrence records

The three studied lineages were modelled in two independent groups: one composed by *S. s. longirostris* (the most divergent mtDNA lineage from all *S. salamandra* lineages), and the other by *S. s. crespoi/morenica.*

Species occurrence records were obtained from our own data, the "Atlas Portugal", "Asociación Herpetológica Española", and other collaborators. Starting with a dataset of 2266 occurrence records of *S. salamandra* covering the Iberian Peninsula, selected points were located within the range of the three studied mtDNA lineages. Posteriorly, a maximum of one presence record per 10 x 10 km grid cell and a minimum distance of 1.5 km between occurrence data (approximately the estimated home range of *Salamandra* species; Schulte *et al.* 2007), were used to avoid potential sampling biases arising from spatial autocorrelation (Fourcade *et al.* 2014). After filtering the occurrence records dataset, a total of 116 occurrence records (26 records for *S. s. longirostris* and 90 *S. s. crespoi/morenica* group) were used (Fig. 11).

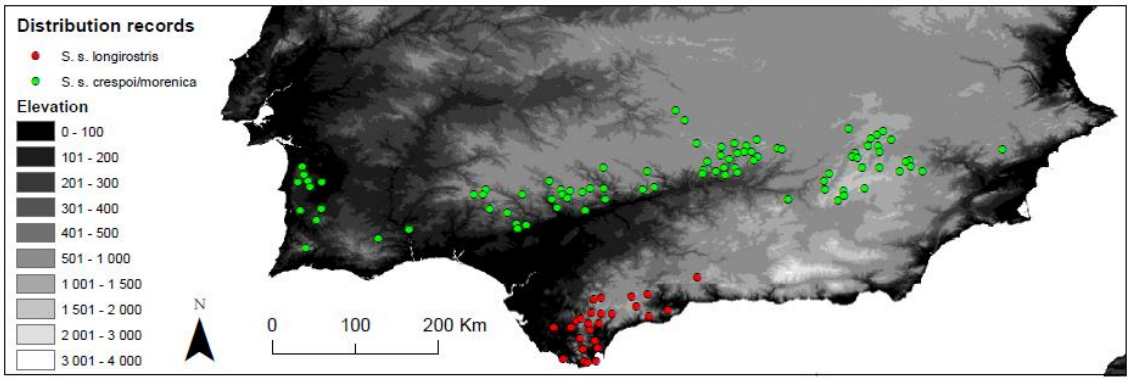


Figure 11 – Digital elevation layer of southern Iberian Peninsula containing occurrence records used for ENMs (green dots - *S. s. crespoi/morenica* mtDNA lineage; red dots - *S. s. longirostris* mtDNA lineage).

2.7.2. Environmental variables

To model the ecological niche for the present in the two groups of mtDNA lineages, 19 bioclimatic layers were downloaded from WorldClim database website (www.worldclim.org) at the resolution of 30 arc-seconds (~1 x 1 km). Two additional non-climatic variables were used for ecological niche modelling – Slope and the Normalized Difference Vegetation Index (NDVI). The former was derived from an altitude layer (~1 x 1 km; WorldClim database website), using the Slope function of ArcGis 10.1 (ESRI). The latter was derived from NDVI layers acquired from the MOD13Q1 product of the MODIS Terra satellite at a resolution of 250 x 250 m, available at NASA's Land Processes Distributed Active Archive Center (LPDAAC, https://lpdaac/usgs.gov; details in supplementary material). To avoid inclusion of correlated variables, a Pearson's correlation test was employed (correlated variables $r^2 \geq 0.70$) in ArcMap (ESRI). If a pair of variables was correlated, then only the variable considered to be biologically significant for *S. salamandra* was included in downstream analyses. The final subset of uncorrelated variables used to build ecological niche models (ENMs) was composed by six climatic variables (BIO3, Isothermality; BIO8, Mean Temperature of Wettest Quarter; BIO9, Mean Temperature of Driest Quarter; BIO11, Mean Temperature of Coldest Quarter; BIO12, Annual Precipitation; BIO18, Precipitation of Warmest Quarter) and 2 non-climatic variables (SLOPE and NDVI) (Table 1).

Table 1 – Set of uncorrelated variables used on ENMs.

Type	Variable	Range and Units
Climatic	Isothermality (BIO3)	3 - 5 °C
	Mean Temperature of Wettest Quarter (BIO8)	-4 - 20 °C
	Mean Temperature of Driest Quarter (BIO9)	1 - 27 °C
	Mean Temperature of Coldest Quarter (BIO11)	-6 - 13 °C
	Annual Precipitation (BIO12)	215 - 1346 mm
	Precipitation of Warmest Quarter (BIO18)	13 - 177 mm
Topographic	Slope	0 - 26 %
Habitat	Normalized Difference Vegetation Index (NDVI)	-2829 – 8708

Modelling for the past was performed only with climatic variables since the remaining variables are not available. The same six climatic variables were downloaded from WorldClim website for the Last Interglacial (LIG; ~120,000 - 140,000 years BP), Last Glacial Maximum (LGM; ~21,000 years BP) and Mid-Holocene (~6,000 years BP) periods. LIG variables were at 30 arc-seconds resolution (~1 x 1 km) and included one Global Circulation Model (GCM): NCAR-CCSM (Community Climate System Model, Otto-Bliesner *et al.* 2008). LGM variables were at 2.5 arc minutes resolution (~5 x 5 km) and accounted for two GCMs: CCSM and MIROC (Model for Interdisciplinary Research on Climate, ver. 3.2, Hasumi & Emori, 2004) from PMIP2. Mid-Holocene variables were

at 2.5 arc minutes resolution (~5 x 5 km) and included one GCM: CCSM (Community Climate System Model, ver. 3, Collins *et al.* 2006) from the Paleoclimate Modelling Intercomparison Project Phase II. All environmental variables and species occurrence records were projected on WGS84 datum.

2.7.3. Ecological Niche-based models

Ecological niche-based models (ENM) for the two main mtDNA lineages (*S. s. longirostris and S. s. crespoi/morenica*) were estimated for the present and the past using a machine learning algorithm based on maximum entropy implemented in MaxEnt 3.3.3k (Phillips *et al.* 2006). MaxEnt predicts areas of environmental suitability based only on two data input: (1) the species presence-only records; and (2) on environmental variables. It is the most widely used ENM algorithm due to its simplicity and accuracy (Fourcade *et al.* 2014), being considered very robust even compared with other more sophisticated methods (Elith *et al.* 2006, 2010).

Species distribution areas for model construction and validation were selected using a buffer of 200 km radius around each lineage occurrence records and projected to the total distribution area. The two models, one for *S. s. longirostris* lineage (26 occurrence records) and other for *S. s. crespoi/morenica* lineage (90 occurrence records), were built for current conditions using 30% of the data records to test the model and 50 replicates of the model for bootstrap analysis, with the remaining settings kept default. The same parameters were used to model the distribution for the past, but the number of replicates of the model for bootstrap analysis was set up to 20 to speed up the analyses, which did not appear to significantly reduce model performance. Model performance was evaluated using the area under the receiver operating characteristic curve (AUC) in Maxent, which ranges from 0 (complete randomness) to 1 (perfect discrimination) (Phillips *et al.* 2006). The resulting ENMs using current environmental data were imported into ArcGIS 10.1 (ESRI) and reclassified into presence-absence using a ten percentile minimum training presence threshold (i.e. 10% of the observations with the lowest probability are considered as absences) given by Maxent. This more conservative threshold was used as an alternative to the minimum presence threshold, to avoid the overestimation of the suitability areas without excluding extensive areas with presence records (Liu *et al.* 2005, 2013). These reclassified ENMs were then displayed together in the same layer to simplify the visualization of the results, and the identification of possible contact zones. ENMs from past conditions were reclassified using the minimum training presence threshold (suitability values higher or equal to the lowest suitability value given to an observation are considered presences). The resulting maps

from current climatic ENMs and the different four past projections, were used to identify the potential suitable areas maintained over time (i.e. stable areas) of each mtDNA lineage. This was done using ArcGIS 10.1 (ESRI) to overlap the five reclassified ENMs projections (Current, Mid-Holocene, LGM (CCSM and MIROC), and LIG) and then extract the common areas between all the suitability maps (i.e. stable areas).

2.7.4. Ecogeographical variables (EGVs) importance

The most important environmental variables explaining the results of each ENM were determined from the average percentage contribution of each variable to the model. Univariate average response curve profiles were generated for the variables identified as the most important for current models, representing how the logistic prediction changes when each variable changes and the others are kept constant.

2.7.5. Niche comparisons

After estimating past and current distributions of the two main mtDNA lineages studied (*S. s. longirostris* and *S. s. crespoi/morenica*), niche differences and niche overlap between them were analyzed. Niche differences were assessed using the "PCA-env" ordination technique (Broennimann *et al.* 2012). To apply this method, the same occurrence records and the eight uncorrelated environmental variables (six climatic and two non-climatic) previously applied in ENM analyses, were employed. This method uses a principal component analysis (PCA) to transform the environmental space of the variables into a two-dimensional surface, which is defined by the first and second principal components. This PCA is calibrated using the environmental data inside minimum convex polygons delimiting occurrence data of each lineage. Niche overlap between the two considered mtDNA lineages was calculated based on the overlap of the occurrence densities using the D metric of Schoener (1968). This metric varies between 0 (no overlap) and 1 (complete overlap). Posteriorly, two statistical tests were performed to test statistically the degree of overlap (i.e. the metric D) between the two niches. First, the equivalency test, pools the occurrences of both group lineages together, extracting randomly two sets of occurrences with the original sample sizes from which it determines overlap scores between them (100 randomizations). This simulated overlap values obtained are compared with the observed overlap. If the value falls beyond 95% of the simulated values, the hypothesis of niche equivalence is rejected. Second, the niche similarity test addresses whether the environmental niche occupied by one lineage can predict occurrences of the other lineage better (or worst) than expected by chance. In other words, it assesses if they are more similar (or different) than expected by chance

by generating 100 randomizations. If the actual overlap between the niches of the different lineages is greater than 95% of the simulated values, then niche similarity is accepted.

2.8. Landscape genetic analyses

To evaluate if present isolation by distance (isolation-by-distance model; IBD; Wright 1943) and isolation by environment (isolation-by-resistance model (IBR) considering IBD influence; McRae 2006) influences the contemporary genetic structure of current populations of *S. s. longirostris* (subset 5), a landscape genetics framework, integrating spatial, environmental and genetic distances was developed. Three matrices were constructed: (1) a matrix of inter-population pairwise genetic distances; (2) a matrix of pairwise euclidean geographic distances; and (3) a matrix of pairwise ecological resistance between demes. In this section, we aimed for a population based approach, using a total of seven populations of *S. s. longirostris*.

Three measures of genetic differentiation were used to construct three distinct matrices of genetic distances: (1) a matrix of pairwise F_{ST}; (2) a matrix of pairwise Jost´s D_{EST}; and (3) a matrix of the conditional genetic distance (cGD) (Dyer *et al.* 2010). The latter metric is based upon the differences in genetic covariation at inter-population level, estimated from population graphs. It implies that populations exchanging more effective migrants will have a higher genetic covariation than populations not exchanging gene flow (genetically more independent). cGD showed to provide more statistical power than pairwise F_{ST} in describing the spatial distribution of genetic variation and therefore, it was included also here. cGD was calculated in R package *gstudio* (Dyer 2014). Additionally, the matrix of Euclidean geographical distances between populations was calculated in GenAlex. Finally, the matrix of pairwise ecological resistance distances was calculated in CIRCUITSCAPE 4.0 (McRae 2006). This software uses an algorithm based on circuit theory. Given that electricity has properties of a random walk in an electric circuit, then resistance parameters can be interpreted as the probability of an individual randomly travelling through the landscape between populations. Unlike least-cost path (Adriaensen *et al.* 2003), the circuit algorithm employed here has the advantage of accounting multiple possible pathways, an interesting feature that is likely more realistic at regional levels (McRae 2006; Spear *et al.* 2010). CIRCUITSCAPE requires as an input a raster layer of resistance and a raster layer depicting only the nodes location (in this case the populations). We reclassified the *S. s. longirotris* ENM probability suitability map obtained in Maxent for present conditions (see above) into a resistance raster layer by a simple inversion of the values (1 – x), and posterior transformation of the resistance

values into percentages (multiplication of resistance values by 100). Input files were analyzed using the pairwise mode. The option to create current maps was used to highlight connective elements in raster frameworks. Other parameters were kept default.

After calculating the pairwise matrices, IBD and IBR matrices were correlated with each of the three calculated genetic distance matrices. First, a Mantel test was conducted to assess significant associations between matrices of pairwise genetic distances (F_{ST}, Jost´s D_{EST}, and cGD) with the matrix of pairwise resistance to gene flow (IBR model) and with the matrix of pairwise geographic Euclidean distances (IBD model). Subsequently, a partial Mantel test was performed. This method correlates two matrices while controlling the covariance of a third matrix (Smouse *et al.* 1986). Thus, if a particular model (IBD or IBR matrix) remained significant after controlling the covariance of a third matrix (IBD or IBR matrix), then the model tested would be acknowledged as significantly affecting genetic differentiation. Both Mantel and partial Mantel tests were performed with 9999 permutations in R package *vegan* (Oksanen 2016).

3. Results

3.1. Phylogenetic analyses

Bayesian analyses of mtDNA sequences (*cyt-b*) identified five well-supported clades (BPP>0.95; Fig. 12, B) corresponding to the five subspecies. *Salamandra s. longirostris* was identified as the sister lineage to a clade that included all the remaining subspecies. This clade was then divided into *S. s. crespoi/morenica* lineage and *S. s. gallaica/bejarae* lineage, that were posteriorly divided into each different clade.

Individuals, and their respective clades were then displayed geographically to delimit the geographic range of each subspecies, allowing us to infer the mtDNA lineage of the sampled demes employed in population genetic analyses (Fig. 12, A).

The nuclear haplotype network (Fig. 12, C) revealed a lack of variability among subspecies, with only five variable positions (three with heterozygous peaks), and no genetic structure or geographic congruence.

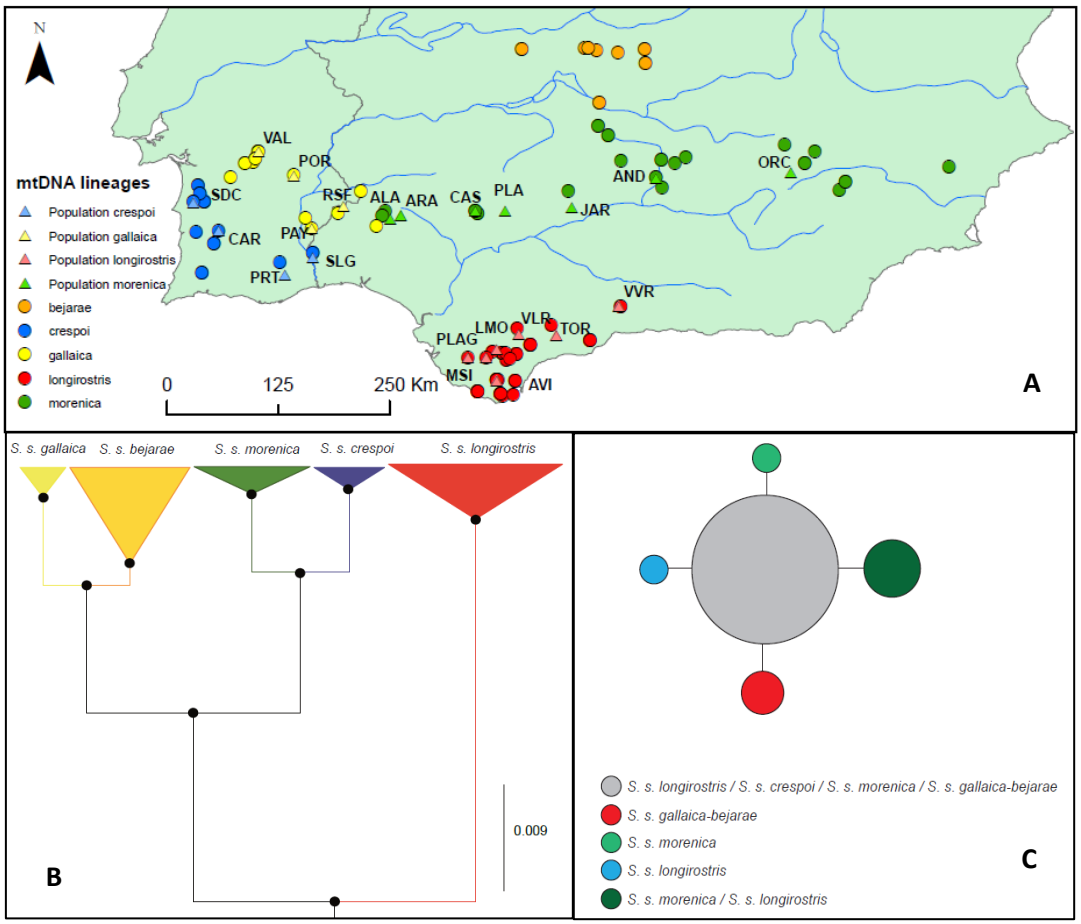


Figure 12. (A) Map displaying the total number of samples used in phylogenetic analyses, and their respective clade identified by color (in circles). Demes used for population genetic analyses, and their attributed mtDNA linegaes identified by color (in triangles). Main rivers represented with blue lines. **(B)** Bayesian phylogenetic tree based on the mtDNA (*cyt-b*) data. Black dotes identify the nodes with posterior probability higher than 0.95. **(C)** *β-fibint7* haplotype network inferred by TCS under the 95% criterion, showing the five different haplotypes. The size of each haplotype symbol is proportional to its frequency, and lines represent mutational steps separating observed haplotypes.

3.2. Microsatellite quality control

From the initial dataset of 535 samples, data filtering procedures (i.e. samples with more than 40% of missing data) reduced the number to a total 460 genotyped individuals. The inclusion of related individuals (39 individuals identified with COLONY) changed little the results of genetic diversity and pairwise differentiation ($r > 0.99$; see next paragraph; results for subset 2 in Table S5 and S6) and therefore, subset 2 was not used in subsequent analyses. Furthermore, a total of 358 individuals had intersecting 2 km radius buffers, being pooled in 22 distinct demes (Fig. 12, A; Table 2) for population-based genetic analyses (subset 1). The subsets 3, 4 and 5 included 15, 15 and seven demes, respectively (Table 2). The remaining 102 individuals were included only in genetic structure analyses.

None microsatellite loci showed evidence of null alleles, large allele dropout or stuttering bands. Eight loci pairs showed signs of linkage disequilibrium (LD), although none of these cases showed a consistent pattern of LD across the studied populations. Additionally, significant deviations from HWE were only found in VVR population due to the presence of four monomorphic loci. Therefore, we opted to maintain all loci for downstream analyses.

3.3. Genetic diversity

Overall, for subset 1, all populations presented high levels of diversity, excepting populations TOR and VVR belonging to *S. s. longirostris* (Table 2). The populations exhibiting the highest values of diversity were SDC, CAR, PRT (of *S. s. crespoi*) and PAY (of *S. s. gallaica*), with Ho ranging from 0.76 to 0.82 and AR from 6.52 to 7.09 in those populations. On the other hand, relatedness values were extremely high in several populations, with populations SLG (0.52; of *S. s. crespoi*), TOR (0.52) and VVR (0.84) (both of *S. s. longirostris*) displaying the highest relatedness (Table 2 see below).

Table 2 - Sampling information and genetic diversity values for each population.

Population	Locality	N	N_A	H_O	H_E	AR	P-AR	R
VAL	Évora, Valverde	21	8.44	0.72	0.75	5.88	0.16	0.27
POR	Évora; Portel	15	6.89	0.72	0.73	5.34	0.21	0.31
PAY	Huelva	14	8.67	0.82	0.81	6.53	0.15	0.14
RSF	Huelva; Rosal de la Frontera	14	6.78	0.76	0.75	5.57	0.51	0.27
SDC	Setúbal; Santiago do Cacém	15	8.67	0.76	0.83	6.52	0.61	0.15
CAR	Beja; S. Martinho da Amoreiras/Carniceiro	19	10.33	0.80	0.85	7.09	0.39	0.14
PRT	Tavira, Portela	15	9.22	0.78	0.79	6.67	0.22	0.23
SLG	Sanlucar de Guadiana	17	4.56	0.69	0.63	3.94	0.31	0.52
ALA	Huelva; Aldea del Cabezuelo, Alájar	14	8.00	0.69	0.82	6.52	0.46	0.11
ARA	Huelva; Aracena	23	10.67	0.71	0.80	6.68	0.20	0.14
CAS	Sevilla; Cazalla de la Sierra	23	6.56	0.75	0.72	5.14	0.01	0.31
PLA	Sevilla; Las Navas de la Concepcion	13	6.00	0.75	0.73	4.90	0.06	0.26
JAR	Córdova, Las Jaras	13	5.44	0.66	0.71	4.56	0.02	0.30
AND	Jaén; Andújar	20	6.33	0.64	0.69	4.79	0.05	0.31
ORC	Jaén; Fte. Peña Alta, Orcera	16	6.22	0.69	0.70	4.92	0.68	0.35
MSI	Cádiz, Medina Siotonia	12	4.67	0.64	0.67	4.04	0.11	0.41
PLAG	Cádiz; Picalho, Sª del Aljibe, Alcalá de los Gazules	19	6.44	0.66	0.72	5.09	0.15	0.33
LMO	Málaga, Laguna del Moral	10	6.33	0.68	0.77	5.72	0.03	0.22
AVI	Cádiz, Arroyo de Valdeinfierno	19	6.89	0.68	0.72	5.18	0.09	0.32
VLR	Cádiz, Villaluenga del Rosario - Llanos del Republicano	15	6.11	0.54	0.65	4.87	0.14	0.41
TOR	Málaga, Camino a la cima del Torrecilla	12	3.89	0.50	0.56	3.47	0.13	0.52
VVR	Málaga, Villanueva del Rosario	19	2.44	0.30	0.28	2.10	0.20	0.84

* Mitochondrial DNA lineages identified with colors. *Salamandra s. gallaica* in yellow, *S. s. crespoi* in blue, *S. s. morenica* in green and *S. s. longirostris* in red.

3.4. Genetic differentiation

Pairwise F_{ST} comparisons between demes were all significant and ranged from 0.022 to 0.565, while pairwise Jost's D_{EST} values were all significant as well and ranged from 0.067 to 0.957. Overall, higher divergence values were found between populations separated by the Guadalquivir river basin (i.e. between *S. s. longirostris* and the other subspecies; Table 3). Populations TOR and VVR presented the highest levels of differentiation among populations of *S. s. longirostris*, especially VVR (Table 3). Among the other subspecies SLG (*S. s. crespoi*) had the highest differentiation values, followed by PRT and CAR (both from *S. s. crespoi*) and ALA (*S. s. morenica*) (Table 3). These patterns were more prominent using Jost's D_{EST} measure.

Results of the first STRUCTURE analysis using the subset 3 depicted K=2 as the K that captures the major genetic structure in the data based on the Evanno criterion ΔK. Although K=2 clearly separates *S. s. longirostris* populations from *S. s. crespoi* and *S. s. morenica* populations (Fig. 13), three individuals appeared with higher degrees of admixture (two individuals in cluster 1 with membership values of 58% and 69% and one in cluster 2 with 77%; Fig. 13). NewHybrids further corroborated these results (no gene flow among groups) since it allowed us to exclude the possibility of the identified admixed individuals being first- or second-generation products of migrants.

With no clear evidence of gene flow among *S. s. longirostris* and the other lineages, further substructure was assessed independently for *S. s. longirostris* (subset 5) and for all the other lineages using the subset 4. With the additional individuals

Table 3. Microsatellite pairwise F_{ST} and Jost's D_{EST} values above and below the diagonal respectively.

F_{ST}/D_{EST}	VAL	POR	SDC	CAR	PRT	PAY	SLG	RSF	ALA	ARA	CAS	PLA	JAR	AND	ORC	MSI	PLAG	LMO	AVI	VLR	TOR	VVR
VAL		0.039	0.061	0.126	0.191	0.082	0.292	0.089	0.151	0.098	0.113	0.146	0.154	0.162	0.128	0.265	0.244	0.210	0.234	0.277	0.309	0.469
POR	0.115		0.072	0.129	0.202	0.100	0.303	0.110	0.167	0.119	0.134	0.166	0.180	0.176	0.158	0.282	0.251	0.221	0.239	0.290	0.312	0.492
SDC	0.237	0.274		0.089	0.136	0.075	0.242	0.113	0.116	0.088	0.101	0.112	0.132	0.136	0.132	0.211	0.184	0.159	0.183	0.239	0.277	0.441
CAR	0.579	0.570	0.516		0.075	0.103	0.169	0.123	0.099	0.088	0.158	0.129	0.147	0.140	0.161	0.191	0.168	0.134	0.173	0.198	0.243	0.402
PRT	0.795	0.816	0.679	0.385		0.122	0.115	0.175	0.126	0.108	0.183	0.142	0.144	0.160	0.193	0.204	0.176	0.144	0.189	0.234	0.287	0.453
PAY	0.313	0.375	0.372	0.577	0.567		0.199	0.086	0.087	0.061	0.063	0.082	0.082	0.097	0.096	0.196	0.191	0.155	0.190	0.230	0.248	0.454
SLG	0.918	0.910	0.839	0.582	0.318	0.627		0.266	0.213	0.199	0.274	0.227	0.241	0.244	0.288	0.296	0.279	0.249	0.289	0.322	0.346	0.529
RSF	0.292	0.349	0.470	0.571	0.717	0.334	0.785		0.139	0.094	0.125	0.145	0.160	0.154	0.140	0.263	0.232	0.200	0.225	0.277	0.312	0.492
ALA	0.640	0.696	0.630	0.576	0.616	0.436	0.686	0.588		0.037	0.147	0.110	0.124	0.139	0.144	0.206	0.188	0.160	0.185	0.219	0.243	0.456
ARA	0.382	0.457	0.433	0.470	0.492	0.279	0.645	0.366	0.172		0.092	0.082	0.082	0.100	0.099	0.219	0.189	0.169	0.195	0.223	0.246	0.417
CAS	0.350	0.403	0.369	0.677	0.683	0.212	0.780	0.384	0.559	0.324		0.045	0.063	0.062	0.097	0.254	0.236	0.215	0.236	0.273	0.316	0.470
PLA	0.494	0.543	0.454	0.580	0.540	0.306	0.614	0.482	0.436	0.308	0.124		0.049	0.036	0.111	0.245	0.216	0.195	0.214	0.266	0.306	0.479
JAR	0.492	0.562	0.508	0.630	0.509	0.284	0.625	0.507	0.468	0.287	0.163	0.132		0.037	0.128	0.271	0.247	0.215	0.231	0.288	0.331	0.488
AND	0.499	0.523	0.488	0.552	0.549	0.320	0.627	0.460	0.496	0.337	0.159	0.093	0.089		0.106	0.255	0.241	0.212	0.235	0.286	0.332	0.469
ORC	0.385	0.465	0.485	0.667	0.697	0.323	0.786	0.417	0.525	0.341	0.259	0.312	0.343	0.270		0.278	0.264	0.234	0.249	0.301	0.342	0.510
MSI	0.900	0.922	0.809	0.783	0.711	0.707	0.761	0.868	0.774	0.823	0.775	0.765	0.822	0.733	0.833		0.111	0.082	0.136	0.208	0.267	0.514
PLAG	0.892	0.879	0.760	0.742	0.662	0.765	0.789	0.828	0.765	0.757	0.783	0.732	0.815	0.758	0.869	0.287		0.022	0.097	0.119	0.245	0.410
LMO	0.853	0.867	0.779	0.704	0.623	0.716	0.753	0.798	0.776	0.790	0.788	0.745	0.785	0.728	0.840	0.232	0.067		0.047	0.109	0.230	0.443
AVI	0.858	0.840	0.772	0.782	0.739	0.772	0.854	0.807	0.767	0.799	0.794	0.736	0.761	0.747	0.815	0.367	0.282	0.146		0.152	0.233	0.400
VLR	0.904	0.910	0.886	0.765	0.801	0.815	0.838	0.889	0.766	0.791	0.815	0.814	0.855	0.820	0.891	0.514	0.299	0.302	0.401		0.299	0.460
TOR	0.894	0.847	0.893	0.829	0.879	0.744	0.781	0.872	0.721	0.761	0.855	0.824	0.871	0.866	0.902	0.592	0.605	0.605	0.569	0.671		0.565
VVR	0.957	0.930	0.897	0.870	0.904	0.915	0.895	0.919	0.881	0.888	0.917	0.859	0.838	0.843	0.939	0.849	0.689	0.746	0.663	0.692	0.865	

* Mitochondrial DNA lineages identified with colors. *Salamandra s. gallaica* in yellow, *S. s. crespoi* in blue, *S. s. morenica* in green and *S. s. longirostris* in red.

sampled to cover gaps between populations, in each subdataset. For subset 5, K=2 was the most supported K in STRUCTURE, according to the Evanno method. Indeed, this K clearly separates the easternmost population VVR from the central group, containing the remaining six *S. s. longirostris* populations. Moreover, the second best (K=3) managed to separate central eastern from central western populations (Fig. 14). Results for the other lineages (subset number 4), revealed K=3 as the most supported K, with populations belonging to different subspecies showing evident patterns of admixture, highlighting existence of gene flow between them (Fig. 15). Two main contact zones could be recognized, one involving the three identified groups, located around the Portuguese-Spanish border (populations PAY, RSF, ALA and ARA; Fig. 15), and another one involving only two of the three identified groups, located in the southwest, specifically in population CAR (Fig. 15).

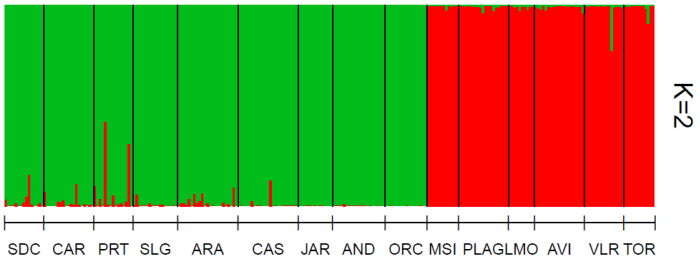


Figure 13. Individual membership to the two main groups inferred by STRUCTURE. *Salamandra s. longirostris* populations in red, and *S. s. crespoi/morenica* population in green. Analysis performed using subset 3.

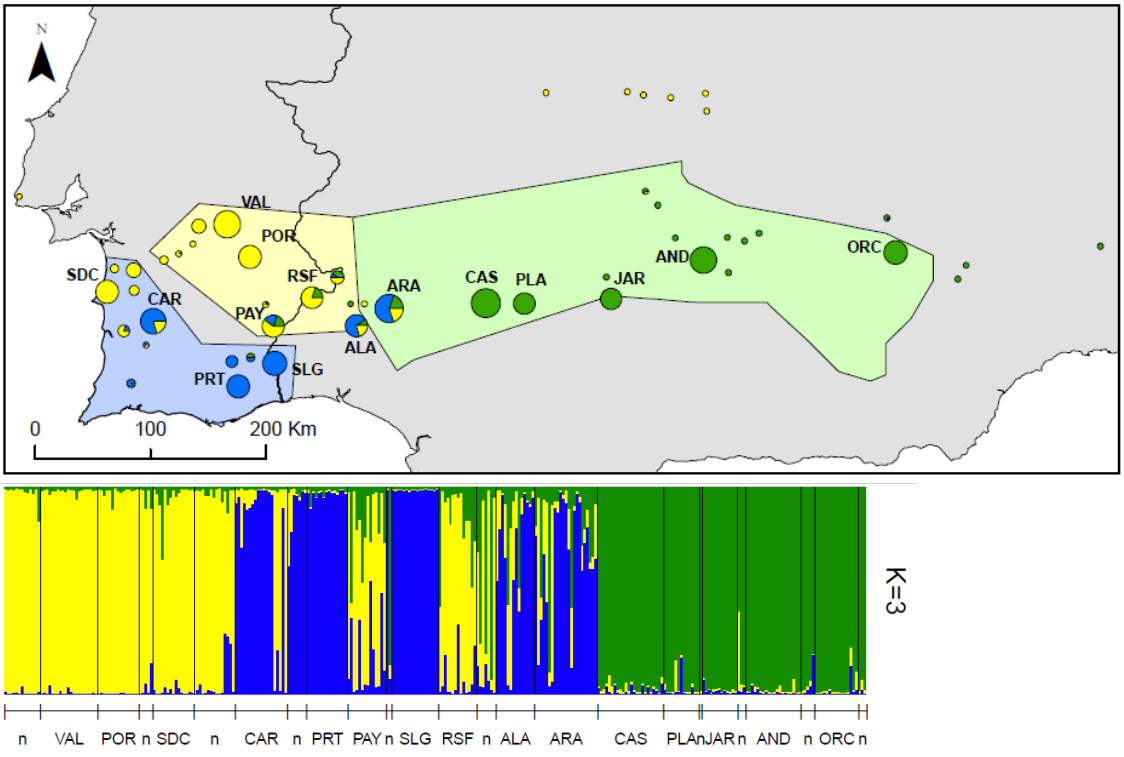


Figure 14. STRUCTURE barplot (below) and map of the study area with population pie charts (above) for the most supported number of clusters (K=3) representing individual cluster membership for populations located north of the Guadalquivir river basin. Individuals without attributed demes, are identified with n. Analysis performed using subset 4, plus scattered samples that were not assigned to any deme. Pie charts size relative to sample size. Colored polygons represent mtDNA lineages distribution, obtained from our data. *Salamandra s. morenica* in green, *S. s. crespoi* in blue

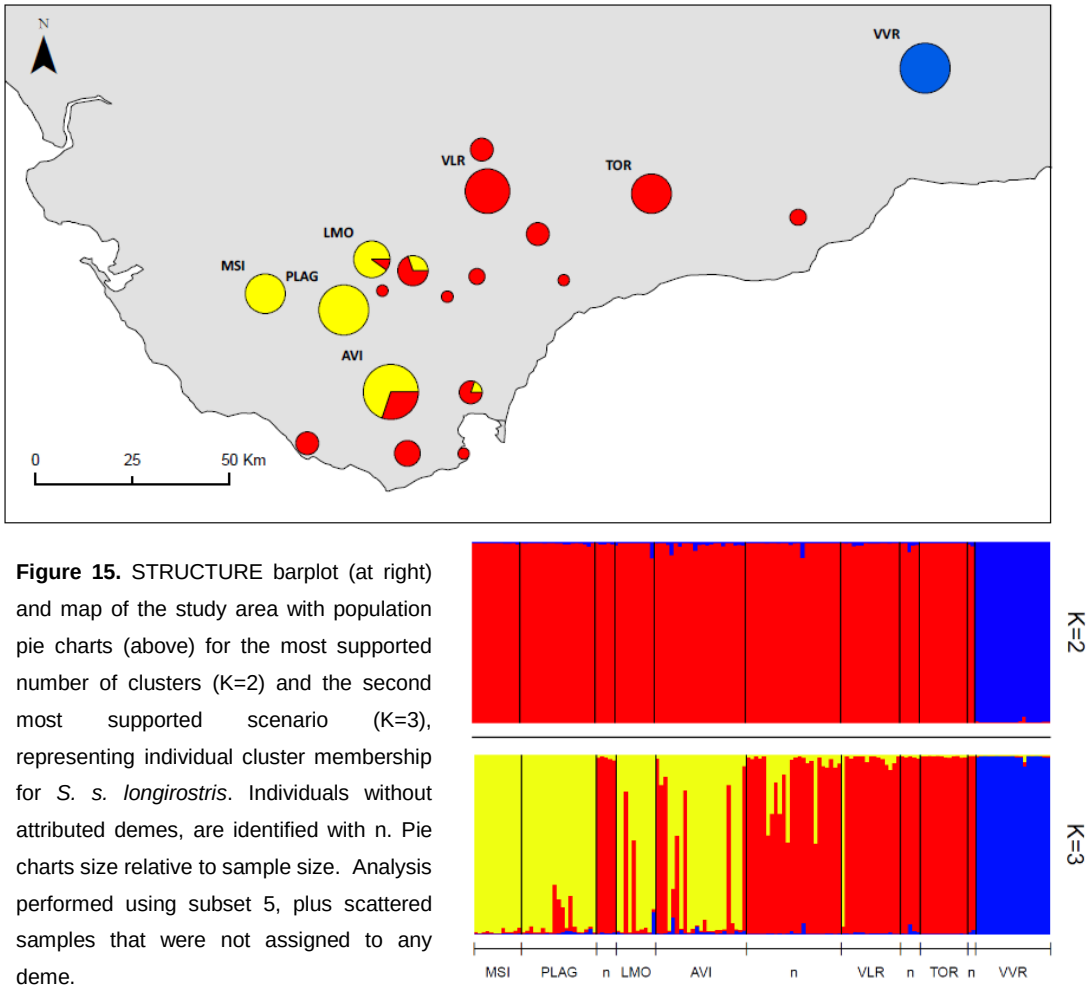


Figure 15. STRUCTURE barplot (at right) and map of the study area with population pie charts (above) for the most supported number of clusters (K=2) and the second most supported scenario (K=3), representing individual cluster membership for *S. s. longirostris*. Individuals without attributed demes, are identified with n. Pie charts size relative to sample size. Analysis performed using subset 5, plus scattered samples that were not assigned to any deme.

3.5 Ecological niche-based modelling

3.5.1. ENMs for the present and the past

All ENMs had a high AUC (AUC > 0.89) suggesting that they have a good predictive power (Table 4). ENMs for *S. s. longirostris* lineage using contemporary environmental variables conformed to the known species distribution, with exception of some specific locations at north of the Guadalquivir river basin that were identified as suitable for the species, more specifically the western Portuguese coast in the extreme north of the study area (Fig. 16). ENMs for *S. s. crespoi/morenica* lineage for the present also conformed to the known species distribution, with the exception of a predicted suitable area located in *S. s. longirostris* distribution range (Fig. 16). By overlapping the layers of the main mtDNA lineages, a distributional gap for *S. salamandra* is evident in the Guadalquivir river basin (Fig. 16).

When comparing the variation on habitat suitability from the past until the present, ENMs revealed distinct patterns between *S. s. longirostris* and *S. s. crespoi/morenica*. For past conditions, ENMs for *S. s. longirostris* varied considerably among the four projected climatic scenarios (LIG, LGM, mid-Holocene and present). Predicting larger and more homogeneous suitable areas for LGM models (CCSM and MIROC), while for other scenarios they were more restricted, especially during the LIG and the current interglacial (Fig. 17). On the other hand, ENMs of *S. s. crespoi/morenica* revealed fewer differences, at least in Spanish locations (i.e. *S. s. morenica* current distribution range; Fig. 17). Southern Portugal (*S. s. crespoi* current distribution range) was identified as an unsuitable are during the LIG (Fig. 17). Through all the scenarios the southern tip of Iberia was identified as an unsuitable area for *S. s. crespoi/morenica* lineage (Fig. 17).

Overlapping maps representing the stable areas for each lineage show no common stable areas between lineages, although contact zones appear as a possibility in some scenarios, mainly during the LGM (Fig. 17 and 18). Additionally, southern Portugal was not recognized as part of the main refugial area for *S. s. crespoi/morenica* lineage, although some really small refugia could be identified (Fig. 17 and 18).

Table 4 - Average (and confidence intervals) training and test area under the curve percent (AUC), and average percentage contribution (and confidence intervals) of each variable, for the two ENMs.

	S. s. longirostris	***S. s. morenica/crespoi***
Training AUC	0.983 (0.979-0.987)	0.919 (0.907-0.931)
Test AUC	0.970 (0.954-0.985)	0.895 (0.874-0.916)
NDVI	22.93 (10.15-35.71)	27.45 (18.06-36.84)
SLOPE	11.21 (2.44-19.97)	17.11 (11.50-22.72)
BIO 12	38.89 (24.94-52.83)	16.65 (10.66-22.64)
BIO 11	5.67 (0.36-10.99)	13.63 (9.19-18.07)
BIO 9	16.43 (10.29-22.57)	6.36 (2.65-10.07)

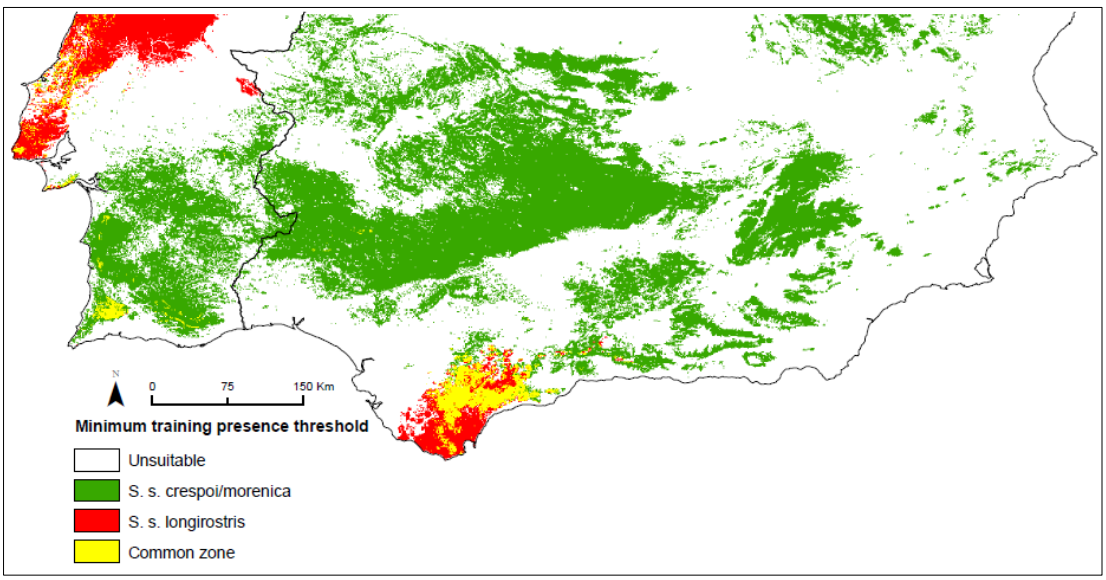


Figure 16. Suitable areas predicted in Maxent for *S. s. longirostris* and *S. s. crespoi/morenica*.

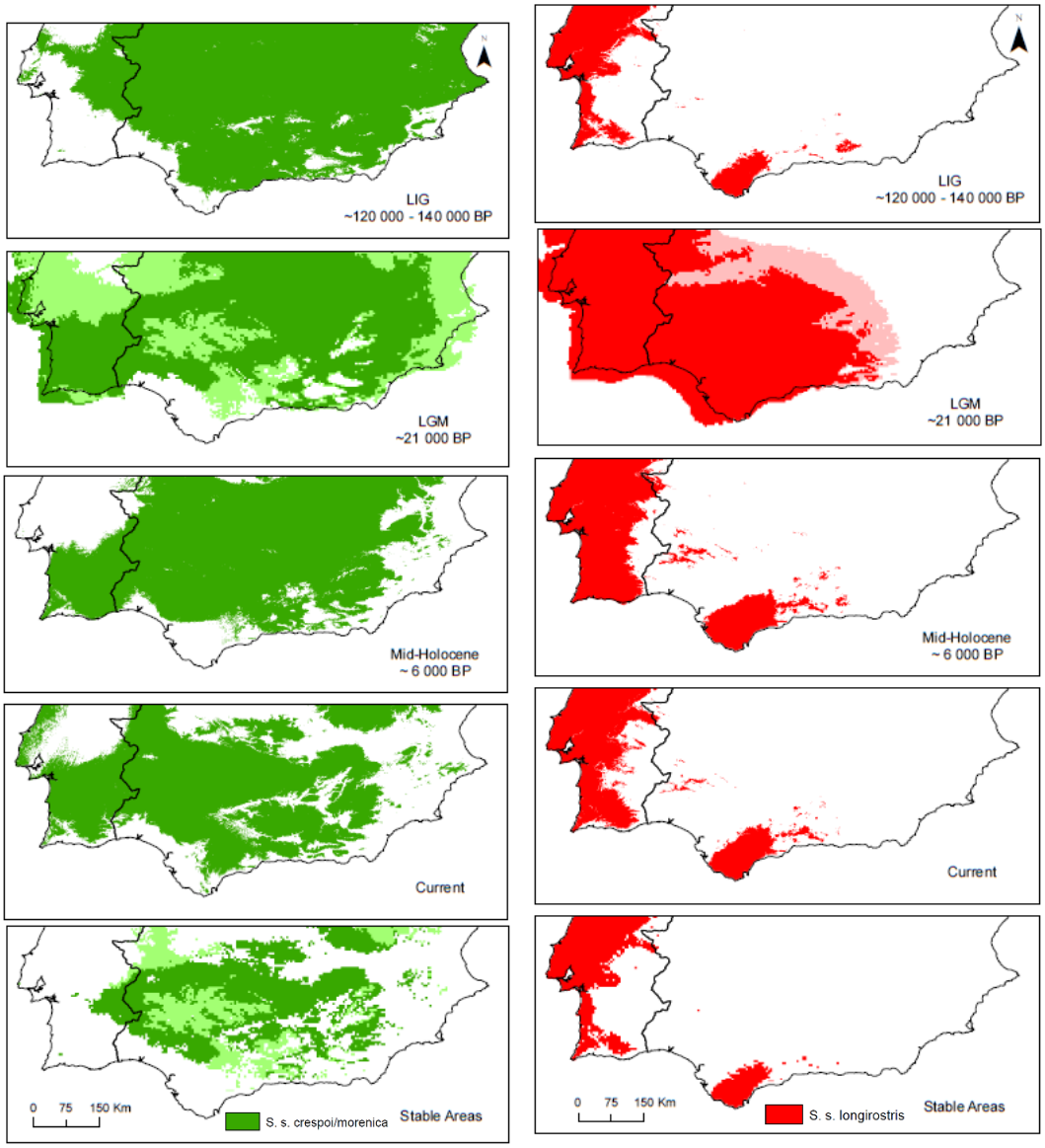


Figure 17. Projection for the Last Interglacial (LIG), Last glacial maximum (LGM), Mid-Holocene, current conditions (Current), and stable climatic areas (Stable Areas) for *S s. longirostris* lineage (red) and *S. s. crespoi/morenica* lineage (green). The two different models used for LGM, are displayed in normal red/green for MIROC and in light red/green for CCSM.

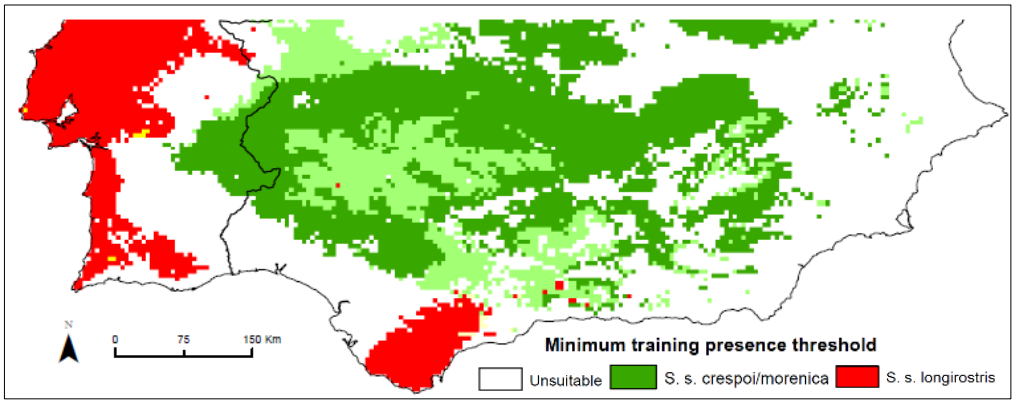


Figure 18. Map overlapping stable climatic areas from the two different lineages. *S. s. longirostris* in red, *S. s. crespoi/morenica* in green, and contact zones in yellow. Lighter green/red represents the suitable area added when using CCSM model for LGM scenario.

3.5.2. Ecogeographical variables (EGVs) importance

The average percent contribution of each variable to the present models revealed that for *S. s. longirostris* lineage the Annual Precipitation (BIO12) and NDVI were the variables with greater predictive power, showing an average percent contribution of 22.93 and 38.89 respectively (Table 4). Whereas for *S. s. crespoi/morenica* lineage the most important variables were NDVI and SLOPE, with an average percent contribution of 27.45 and 17.11 respectively (Table 4).

Response curves showed overall similar patterns in both lineages (Fig. 19). With both lineages presenting higher probability of occurrence for higher NDVI values, whereas low precipitation and slope were associated with a lower probability of occurrence (Fig. 19).

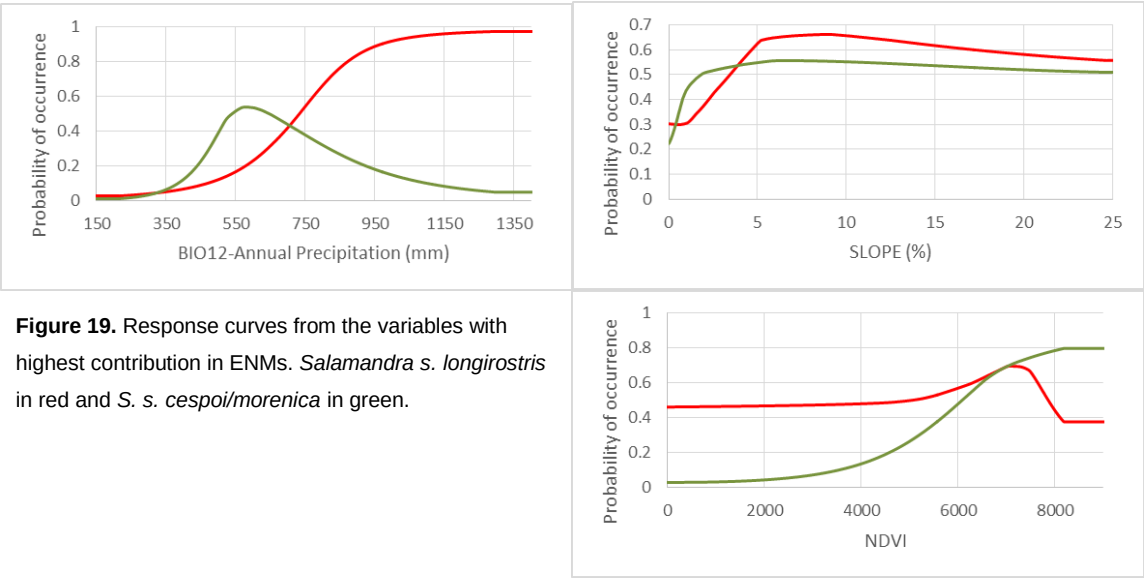


Figure 19. Response curves from the variables with highest contribution in ENMs. *Salamandra s. longirostris* in red and *S. s. cespoi/morenica* in green.

3.5.3. Niche comparisons

PCA-env niche overlap analyses revealed a small niche overlap (D = 0.164) between both studied lineages. Niche equivalency was rejected ($p<0.05$) and thus, niches are not statistically equivalent (Fig. 20). Niche similarity tests indicate that *S. s. crespoi/morenica* niche predicted *S. s. longirostris* (X → Y) occurrence data better than expected by chance ($p<0.05$), while in the other direction of the test (Y → X) the hypothesis of retained niche similarity cannot be rejected ($p>0.05$), meaning that the measured niche overlap between different lineages can be explained by the available environmental space of each lineage, instead of their possible niche divergence (Fig. 20).

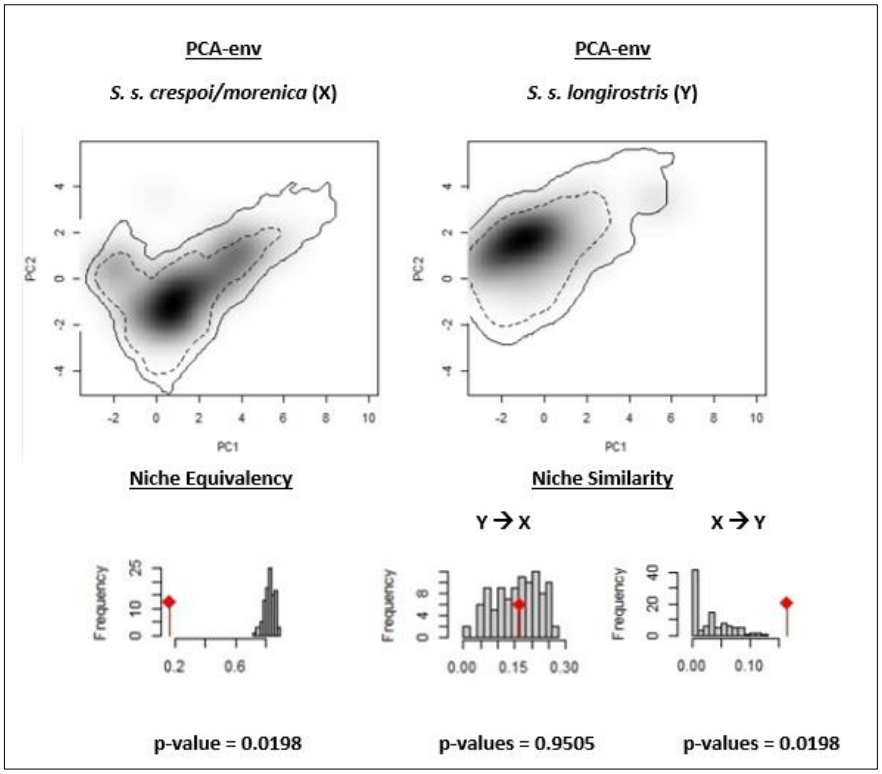


Figure 20. Niches of *S. s. crespoi/morenica* and *S. s. longirostris* in a 2D environmental space. The solid and dashed contour lines illustrate 100% and 50% of the background environment, respectively. Panels in the bottom display the null distributions from niche equivalency and similarity tests (100 randomizations each) with the red dot being correspondent to the actual niche overlap.

3.6. Landscape Genetics

The environmental resistance map yielded by CIRCUITSCAPE revealed overall low resistance to gene flow among populations of *S. s. longirostris*, with the exception of population VVR (Fig. 21). This population was isolated in the easternmost extent of *S. s. longirostris* range, ca. 50 Km away from the closest population, by an area with extremely low connectivity values (Fig. 21). Mantel tests evaluating IBD and IBR models revealed significant ($p<0.05$) and positive correlation values (from 0.678 to 0.870) between geographical/resistance distances and all tested genetic differentiation measures for (F_{ST}, Jost´s D_{EST}, and cGD) (Table 5). When the covariance of IBR matrix was controlled using the IBD matrix and vice-versa, partial Mantel tests presented non-significant correlations ($p>0.05$) for both models, although correlation values for IBR models were higher for pairwise F_{ST} and Jost´s D_{EST} (Table 5). IBD models presented the inverse pattern, with higher correlation values for pairwise cGD and lower for F_{ST} and Jost´s D_{EST} (Table 5).

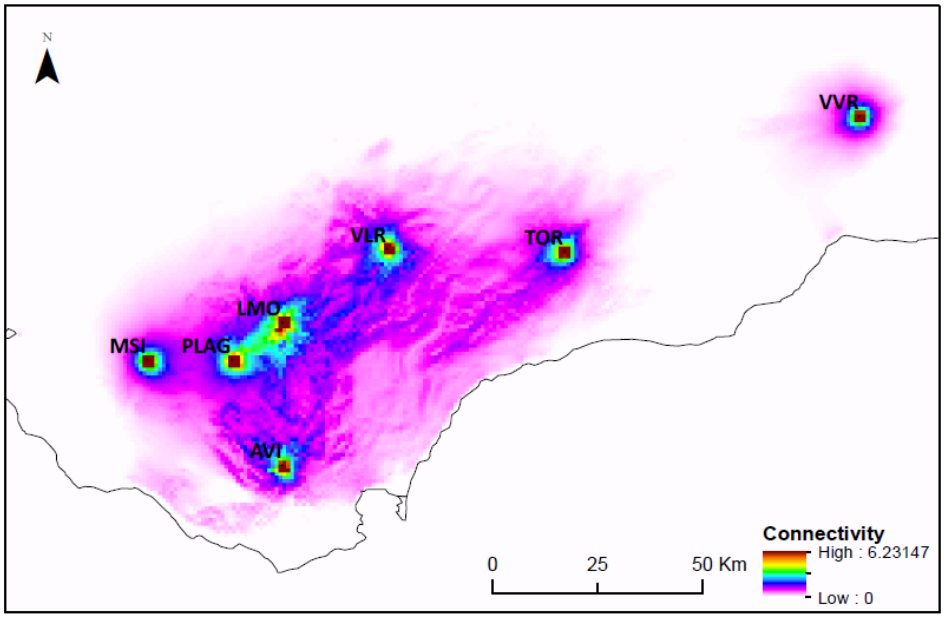


Figure 21. Current maps obtained from CIRCUITSCAPE highlighting connectivity among populations.

Table 5 – Results from Mantel and Partial Mantel tests. Isolation by distance models using geographic Euclidean distances matrix; IBR models using gene flow resistance matrix; IBD/IBR meaning IBD when controlling for IBR and vice-versa.

Test	Model	Genetic distance	r	p-value
Mantel Test	IBD	Fst	**0.815**	**0.004**
		Jost´s D	**0.839**	**< 0.001**
		cGD	**0.775**	**< 0.001**
	IBR	Fst	**0.870**	**0.018**
		Jost´s D	**0.861**	**0.003**
		cGD	**0.678**	**0.003**
Partial Mantel Test	IBD/IBR	Fst	0.049	0.462
		Jost´s D	0.215	0.197
		cGD	0.531	0.052
	IBR/IBD	Fst	0.524	0.067
		Jost´s D	0.408	0.164
		cGD	-0.170	0.597

4. Discussion

4.1. Geographic distribution of *S. salamandra* mtDNA lineages in southern Iberian Peninsula

Our phylogenetic analyses based on mitochondrial DNA sequences (*cyt-b*) identified five different lineages, associated to the five subspecies occurring in the southern Iberia (Fig. 12, B; *S. s. gallaica*, *S. s. bejarae*, *S. s. crespoi, S. s. morenica, and S. s longirostris*). In congruence with previous studies (García-París *et al.* 1998; 2003; Vences *et al.* 2014; Beukema *et al.* 2016), *S. s. longirostris* was identified as the sister lineage to a clade including all the other subspecies. This clade was then divided into one lineage comprising *S. s. gallaica/bejarae* and another with *S. s. crespoi/morenica* (Fig. 12, B).

The geographic representation of these mtDNA lineages, was also congruent with previous studies (García-París *et al.* 1998; Steinfartz *et al.* 2000; García-París *et al.* 2003; Vences *et al.* 2014), with *S. s. crespoi* and *S. s. morenica* showing a relatively continuous distribution along a west-east axis from Alentejo in Portugal to Murcia in Spain, with a secondary contact zone near the Portuguese-Spanish border. *Salamandra s. gallaica* was also identified in this contact zone, which is the southernmost extent of its distribution range. Individuals from the mountain range of Toledo (north of *S. s. morenica* range) belong to *S. s. bejarae* as expected. (Fig. 12, A). *Salamandra s. longirostris* was geographically isolated from other taxa, south of the Guadalquivir river basin. (Fig. 12, A).

4.2. Evolutionary history and climatic niche of southern *Salamandra* subspecies

Our ecological niche models for the present identified a distributional gap between the two mitochondrial lineages (*S. s. longirostris* and *S. s. crespoi/morenica*; Fig. 16). These results support the current allopatric isolation of *S. s. longirostris*, and point Guadalquivir river basin as a current barrier to *S. salamandra* dispersal. Moreover, ENMs for *S. s. crespoi/morenica* identified an area with lower suitability values around the Portuguese-Spanish border, which represents the main contact zone between these two subspecies.

Identification of the variables with greater predictive power (NDVI, BIO12, and SLOPE), identified vegetation, precipitation and slope as the ecological factors playing an important role in the distribution of *S. salamandra*, as expected by its life-history traits (i.e. terrestrial species associated with humid and shaded environments, with a water

dependent larval phase). Low levels of precipitation and slope, appeared to affect negatively this species, leading to low probabilities of occurrence, while higher levels of vegetation seemed to promote species occurrence (Fig. 19). The Guadalquivir river basin is widely used for agriculture activities, resulting in a region extremely affected by habitat destruction and increasing aridity, which probably intensifies its role as barrier to *S. salamandra* dispersal (Pleguezuelos 2004).

Speciation processes are highly linked with species ecological niches (Wiens 2004). Niche conservatism for example is linked to allopatric speciation events, resulting in genetic divergence between populations for which gene flow has been disrupted due to unsuitable environments (Wiens & Graham 2005). Niche equivalency analyses of the two different lineages (*S. s. longirostris* and *S. s. crespoi/morenica*) identified their niches as non-equivalent. These results could be explained by differences in their available habitat, and not necessarily as the result of species niche divergence, as it was suggested by niche similarity analysis that revealed some degree of niche conservatism between lineages. Thus, we cannot identify a specific scenario for niche evolution, and leave open the possibilities of an allopatric divergence process resulting from niche conservatism or niche divergence.

Past (LIG, LGM and Mid-Holocene) and present projections of climatic ENMs revealed isolated stable areas for each lineage (*S. s. longirostris* and *S. s. crespoi/morenica*; Fig. 18), supporting the existence of multiple isolated climatic refugia in southern Iberia for *S. salamandra*. Potential contact zones, resulting from range changes through the different scenarios, could also be identified, particularly during LGM (Fig. 17). This results are both congruent with the existence of multiple isolated refugia in southern Iberia during Quaternary climatic fluctuations. During these climatic fluctuations *S. salamandra* would track its suitable climatic conditions by expanding or contracting its range to different latitudes or altitudes, and thus range contractions could potentially lead to processes of allopatric subspeciation, while posterior range expansions resulted in secondary contact zones (e.g. Garca-Paris *et al.* 2003; Pereira *et al.* 2016). Additionally, the existence of climatic stable areas has been linked to allopatric speciation events and concordant genetic patterns for other Iberian taxa (e.g. Gonçalves *et al.* 2009; Santos *et al.* 2012; Velo-Antón *et al.* 2012; Abellán & Svenning 2014; Martinez-Freiria et al., 2015).

4.3. Introgression or incomplete lineage sorting?

Analysis of nuclear data revealed a lack of variability between subspecies (Fig. 12, C), contrasting with the deep phylogenetic divergence found among mitochondrial lineages (Fig. 12, B), especially in the case of *S. s. longirostris* (Fig. 12). Such cases of mitonuclear discordances are common in animal species (Toews & Brelsford 2012), and can be explained by numerous reasons. Here we discuss a possible case of incomplete lineage sorting (ILS) or an introgression event as the main potential drivers for this mitonuclear discordance patterns. Incomplete lineage sorting results from stochastic nature of the coalescent process, with each marker possibly having its own phylogenetic tree, associated with their characteristics or just random chance. For instance, the mitonuclear discordance pattern observed in our data could result from the longer persistence of ancestral nuclear polymorphisms comparing to ancestral mitochondrial polymorphisms, resulting essentially from the uniparentelly inheritance of the mitochondrial genome. Nuclear introgression among subspecies could also result in the replacement of nuclear genomes, while historical mitochondrial lineages are still preserved. This nuclear introgression events can be associated to numerous evolutionary scenarios such as differential selection on mitochondrial and nuclear genomes, or biased movement of either marker driven by demographic asymmetries (e.g. sex-biased dispersal) (Toews & Brelsford 2012).

Such cases of mitonuclear discordances have been identified in *S. salamandra* studies (García-París *et al.* 2003; Pereira *et al.* 2016), both studies identified introgression events associated to range expansions as the most likely scenario for their mitonuclear discordance patterns. Actually, most taxa displaying patterns of biogeographic mitonuclear discordance correspond to groups that were isolated for long periods of time, and are either currently in secondary-contact zones or have experienced secondary-contact events at some point in the past (Toews & Brelsford 2012). This scenario of isolated populations followed by secondary-contact events during range expansions is the expected scenario for species inhabiting in southern Iberia refugia during the past climatic fluctuations, as suggested by our ENMs. Considering this, we point to introgression as the most likely cause of the observed mitonuclear discordance. However, we cannot rule out the potential case of ILS, or identify the underlying mechanism that resulted in this introgression patterns. Still, we can state several possibilities that would result in the same genetic patterns observed here, such as sex-biased dispersal, mitochondrial selection, or even a purely neutral/random introgression event where the native mitochondrial lineage is captured by the invading nuclear genome

due to demographic effects during a range expansion, favored by mtDNA smaller effective population size (Excoffier *et al.* 2009).

4.4. Gene flow between *S. salamandra* subspecies

Microsatellite analyses revealed current genetic isolation of *S. s. longirostris*, showing high levels of genetic differentiation and absence of gene flow among populations of other subspecies (F_{ST}= 0.134 – 0.510; Table 4). These results, together with our present ENMs, support the role of Guadalquivir river basin as a barrier to gene flow, which seems to be playing an important role in species differentiation in southern Iberian *species* (e.g. García-París *et al.* 1998: Velo-Antón et al. 2012).

The genetic structure detected among subspecies located north of the Guadalquivir river basin (*S. s. gallaica*, *S. s. bejarae S. s. crespoi, S. s. morenica*), recovered slightly different geographic patterns from the mtDNA data, with detected admixture between the three subspecies, along two different contact zones. The first contact zone involve gene flow among the three subspecies (*S. s. gallaica*, *S. s. crespoi* and *S. s. morenica*) around the Portuguese-Spanish border (Fig. 14), along ca. 100 Km West-East axis, going from Paymogo in Huelva (*S. s. gallaica* mtDNA lineage location), to Aracena in Huelva (*S. s. morenica* mtDNA lineage location). The second contact zone was located in the extreme north of *S. s. crespoi* distribution, and involved gene flow only among *S. s. crespoi* and *S. s. gallaica* (Fig. 14).

These results present again a case of mitonuclear biogeographic discordance. This pattern was characterized by clear mtDNA distribution of each subspecies, while a more diffuse pattern was obtained with microsatellites, supporting current nuclear gene flow among mtDNA lineages across the contact zones. The most evident pattern of gene flow involves *S. s. crespoi* nuclear introgression into *S. s. morenica* (in the contact zone around the Portuguese-Spanish border), and also nuclear introgression of *S. s. gallaica* in *S. s. crespoi* in the second contact zone. Even though we cannot accurately identify the mechanism resulting in this mitonuclear discordance pattern, we suggest the hypothesis of female phylopatry and male dispersal as the most likely scenario. This sex-biased hypothesis was already documented in the closely related *Salamandra atra* (Helfer *et al.* 2012), and found to be consistent with our scenario of introgression proposed in the previous chapter, as with other scenarios of introgression in *S. salamandra* (Garcia-Paris *et al.* 2003; Pereira *et al.* 2016). This specific scenario would imply *S. s. crespoi* male dispersal in a west to east direction, across the Portuguese-Spanish border, and posterior hybridization with *S. s. morenica.* While, in the other

contact zone, males from *S. s. gallaica* would disperse to the south, and mate with *S. s. crespoi*.

4.5. *Salamandra salamandra longirostris* habitat connectivity

Contemporary patterns of genetic structure within *S. s. longirostris* identified one isolated population in the easternmost extent of its range (population VVR; Fig. 15). This, together with the high relatedness values (r=0.84) obtained for this population, suggest a demographic scenario with extreme levels of inbreeding followed by genetic drift to explain its current genetic isolation. The observation of adults is extremely rare in this region, which could result in sampling of related larvae, and thus leading to similar genetic patterns. Therefore we cannot rule out this possibility, even with the maintenance of extremely high relatedness values after removing potential full-siblings or parent-offspring (see results).

Structure analyses also identified a subtle genetic structure in the remaining populations, roughly separating the central group in central western and central eastern populations (Fig. 15), with the Betic chain possibly influencing gene flow between western and eastern populations.

The resistance matrix showing the influence of habitat and climate on *S. s. longirostris* dispersal revealed overall low resistance to dispersal among populations, with the exception of population VVR (Fig. 21). This agrees with the results of genetic diversity and differentiation analyses that displayed the lowest diversity and the highest differentiation values for population VVR. This highlights the lack of connectivity between VVR population and all the other *S. s. longirostris* populations, in agreement with our ENM (Fig. 16) and ENMs developed by Romero (2013) and suggesting that populations in the eastern most extent of *S. s. longirostris* range are extremely isolated, and possibly susceptible to inbreeding events.

Isolation models revealed that the present genetic structure of *S. s. longirostris* populations were influenced by distance alone (IBD), and by distance together with environmental characteristics (IBR). Furthermore, if we accept the conditional genetic distance (cGD) as the most appropriate genetic measure for landscape genetic analysis, as suggested by Dyer (2010), we can assume that geographic distance is the factor with the highest influence on the genetic structure of *S. s. longirostris* populations. However, these results might be biased by the sampling distribution of *S. s. longirostris* populations due to the large distribution gap occurring between VVR and all the other populations, associated with extreme genetic values of diversity and differentiation of this population.

Isolation by environment analyses, which differentiate the influence of habitat and/or climatic heterogeneity from the influence of isolation by distance on contemporary genetic structure patterns, were also performed by controlling the effects of distance in the IBR model. The obtained results were all non-significant (Table 6), therefore no further inferences on a potential role of the environment in *S. s. longirostris* genetic structure could be made. Nonetheless, the resistance model based on the ecological niche of the subspecies was not a good predictor of its genetic structure, which was better explained by a simple isolation by distance model. However, this identification of geographic distance as the main driver of genetic differentiation among *S. s. longirostris* populations is probably a byproduct of its geographic distribution, with VVR as the more genetically and geographically distant population, and all the remaining populations roughly grouped in the same region. To avoid potential bias in the analyses, further analyses assessing habitat connectivity between *S. s. longirostris* populations should be focused in the relatively grouped populations, excluding the distant eastern population (VVR).

4.6. Taxonomic implications

The potential elevation of *S. s. longirostris* to full species status has been in debate for some years now. With some authors in favor (Dubois and Raffaëlli 2009; considering its allopatric distribution, genetic divergence at mtDNA level and specific morphology), and others against, suggesting the maintenance of the subspecies status, due to conflicting phylogenetic trees, and the absence of studies addressing gene flow between *S. s. longirostris* and the neighboring subspecies (Speybroeck *et al.* 2010).

In this integrative study, using genetic and ecological data, we identified *S. s. longirostris* as an isolated unit, with the Guadalquivir river basin recognized as the main barrier to current gene flow, probably with an accentuated effect due to anthropogenic land-use changes. These results, together with previous studies, allow us to recognize *S. s. longirostris* as a current genetically and geographically isolated subspecies. Moreover, the identification of its particular niche (i.e. non-identical to the niches of neighboring lineages), opens a possibility to a past or recently adaptive divergence process in *S. s. longirostris* (Wiens 2004). However, owing to mitonuclear discordance patterns (*cyt-b* and *β-fibint7*), we cannot accurately identify for how long *S. s. longirostris* has been an independent evolving lineage, and we can question its distinct potential (e.g. different phenotype) in comparison with other *S. salamandra* subspecies. Considering this, future studies should carefully evaluate the phenotypic distinctiveness of *S. s. longirostris*, and ultimately determine its taxonomic status.

4.7. Evolutionary history of southern Iberia *S. salamandra* subspecies

Here, together with previous studies and our own results, we recreate a possible evolutionary scenario for southern Iberia *S. salamandra* species. While doing so we enumerate our main conclusions.

Genetic patterns observed in this study can be explained by recent historical subdivisions (Pliocene/Pleistocene) of these five *S. salamandra* subspecies in southern Iberia. This scenario is compatible with the hypothesis of multiple isolated refugia in southern Iberia (Gomez & Lunt 2007; Centeno-Cuadros et al. 2009; Gonçalves et al. 2009; Martínez-Solano et al. 2006; Miraldo et al. 2011; Velo-Antón et al. 2012) , which is supported by our past projections of ENMs that identified two main isolated climatic refugia, and potential contact zones events in the past as well. The range contractions and expansions, associated with this scenario, could potentially result in cases of allopatric isolation (associated with differentiation processes) and posteriorly secondary-contact events (allowing gene flow). These secondary-contact events could eventually result in nuclear introgression patterns (Excoffier *et al.* 2009; Petit & Excoffier 2009), which we point as the most likely scenario, considering previous studies on *S. salamandra* that detected evidence for introgression events (Garcia-Paris *et al.* 2003; Pereira *et al.* 2016). Furthermore, we discuss the possibility of this introgression pattern being the result of male biased dispersal in *Salamandra*. Regarding the identification of this mechanism in the closely related *Salamandra atra* (Helfer *et al.* 2012), and with our mitonuclear discordance patterns in the two current contact zones, suggesting this mechanism as the main driver of introgression. However, we cannot exclude other scenarios such as purely neutral introgression events caused by range expansions (Excoffier *et al.* 2009), mitochondrial selection (Welch *et al.* 2014), or even ILS (Toews & Brelsford 2012) – hypothesis that should be further tested in future investigations.

5. References

Abellán P, Svenning J-C (2014) Refugia within refugia–patterns in endemism and genetic divergence are linked to Late Quaternary climate stability in the Iberian Peninsula. *Biological Journal of the Linnean Society*, **113**, 13–28.

Adriaensen F, Chardon JP, De Blust G *et al.* (2003) The application of “least-cost” modelling as a functional landscape model. *Landscape and Urban Planning*, **64**, 233–247.

Alvarado-Serrano DF, Knowles LL (2014) Ecological niche models in phylogeographic studies: Applications, advances and precautions. *Molecular Ecology Resources*, **14**, 233–248.

Álvarez D, Lourenço A, Oro D, Velo-Antón G (2015) Assessment of census (N) and effective population size (Ne) reveals consistency of Ne single-sample estimators and a high Ne / N ratio in an urban and isolated population of fire salamanders. *Concervation genetic resources*.

Anderson EC, Thompson EA (2002) A Model-Based Method for Identifying Species Hybrids Using Multilocus Genetic Data. *Genetics Society of America*, **1229**, 1217–1229.

Arntzen JW, García-París M (1995) Morphological and allozyme studies of midwife toads (genus *Alytes*), including the description of two new taxa from Spain. *Contributions to Zoology*, **65**, 5–34.

Balkenhol N, Waits L, Cushman S, Storfer A (2016) *Landscape Genetics: Concepts, Methods, Applications*. Wiley Blackwell.

Barbosa AM, Sillero N, Martínez-Freiría F, Real R (2012) ECOLOGICAL NICHE MODELS IN MEDITERRANEAN HERPETOLOGY: PAST, PRESENT AND FUTURE. In: *Ecological Niche Models in Mediterranean Herpetology*, pp. 173–204.

Barluenga M, Stölting KN, Salzburger W, Muschick M, Meyer A (2006) Sympatric speciation in Nicaraguan crater lake cichlid fish. *Nature*, **439**, 719–723.

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society B*, **57**, 289–300.

Beukema W, Nicieza AG, Lourenço A, Velo-Antón G (2016) Colour polymorphism in Salamandra salamandra (Amphibia: Urodela), revealed by a lack of genetic and environmental differentiation between distinct phenotypes. *Journal of Zoological Systematics and Evolutionary Research*, 1–10.

Broennimann O, Fitzpatrick MC, Pearman PB *et al.* (2012) Measuring ecological niche overlap from occurrence and spatial environmental data. *Global Ecology and*

Biogeography, **21**, 481–497.

Butlin R, Debelle A, Kerth C *et al.* (2012) What do we need to know about speciation? *Trends in Ecology and Evolution*, **27**, 27–39.

Centeno-Cuadros A, Delibes M, Godoy JA (2009) Phylogeography of Southern Water Vole (*Arvicola sapidus*): Evidence for refugia within the Iberian glacial refugium? *Molecular Ecology*, **18**, 3652–3667.

Chybicki IJ, Burczyk J (2009) Simultaneous estimation of null alleles and inbreeding coefficients. *Journal of Heredity*, **100**, 106–113.

Clement M, Posada D, Crandall KA (2000) TCS: A computer program to estimate gene genealogies. *Molecular Ecology*, **9**, 1657–1659.

Collins WD, Bitz CM, Blackmon ML *et al.* (2006) The Community Climate System Model version 3 (CCSM3). *Journal of Climate*, **19**, 2122–2143.

Coyne JA and HAO (2004) Species: Reality and Concepts. In: *Speciation*, pp. 9–54. Sinauer Associate.

Cushman SA (2006) Effects of habitat loss and fragmentation on amphibians: A review and prospectus. *Biological Conservation*, **128**, 231–240.

Darriba D, Taboada GL, Doallo R, Posada D (2012) jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods*, **9**, 772–772.

Dent A. E, VonHoldt BM (2012) STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, **4**, 359–361.

Donaire-Barroso D, Comas M del M, Carranza S, Escoriza D (2006) Rediscovery of *Salamandra algira* Bedriaga, 1833 from the Beni Snassen massif (Morocco) and phylogenetic relationships of North African Salamandra. *Amphibia-Reptilia*, **27**, 448–455.

Drummond AJ, Suchard MA, Xie D, Rambaut A (2012) Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution*, **29**, 1969–1973.

Dyer RJ, Nason JD, Garrick RC (2010) Landscape modelling of gene flow: Improved power using conditional genetic distance derived from the topology of population networks. *Molecular Ecology*, **19**, 3746–3759.

Elith J, Graham CH, Anderson RP *et al.* (2006) Novel methods improve prediction of species distributions from occurrence data. *Ecography*, **29**, 129–151.

Elith J, Kearney M, Phillips S (2010) The art of modelling range-shifting species. *Methods in Ecology and Evolution*, **1**, 330–342.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Molecular Ecology*, **14**, 2611–2620.

Excoffier L, Foll M, Petit RJ (2009) Genetic Consequences of Range Expansions. *Annual Review of Ecology, Evolution, and Systematics*, **40**, 481–501.

Fisher-Reid MC, Engstrom TN, Kuczynski CA, Stephens PR, Wiens JJ (2013) Parapatric divergence of sympatric morphs in a salamander: Incipient speciation on Long Island? *Molecular Ecology*, **22**, 4681–4694.

Fourcade Y, Engler JO, Rödder D, Secondi J (2014) Mapping species distributions with MAXENT using a geographically biased sample of presence data: A performance assessment of methods for correcting sampling bias. *PLoS ONE*, **9**, 1–13.

Francis RM (2016) pophelper: An r package and web app to analyse and visualize population structure. *Molecular Ecology Resources*.

Futuyma DJ (2013) Speciation. In: *Evolution*, pp. 483–512.

García-París M, Alcobendas M, Alberch P (1998) Influence of the Gualdaquivir River Basin on mitochondrial DNA evolution of Salamandra salamandra (Caudata: Salamandridae) from southern Spain. *Copeia*, **1998**, 173–176.

Garcia-Paris M, Alcobendas M, Buckley D, Wake DB (2003) Dispersal Of Viviparity Across Contact Zones In Iberian Populations Of Fire Salamanders (*Salamandra*) Inferred From Discordance Of Genetic And Morphological Traits. *Evolution*, **57**, 129–143.

Goldberg CS, Waits LP (2010) Quantification and reduction of bias from sampling larvae to infer population and landscape genetic structure. *Molecular Ecology Resources*, **10**, 304–313.

Gomez A, Lunt DH (2007) Refugia within refugia: patterns ofphylogeographic concordance in Iberian Peninsula. In: *Phylogeography of southern european refugia*, pp. 155–188.

Gonçalves H, Martínez-Solano I, Pereira RJ *et al.* (2009) High levels of population subdivision in a morphologically conserved Mediterranean toad (*Alytes cisternasii*) result from recent, multiple refugia: Evidence from mtDNA, microsatellites and nuclear genealogies. *Molecular Ecology*, **18**, 5143–5160.

Greven H (2011) Maternal Adaptations to Reproductive Modes in Amphibians.

Habel JC, Zachos FE, Dapporto L *et al.* (2015) Population genetics revisited - towards a multidisciplinary research field. *Biological Journal of the Linnean Society*, **115**, 1–12.

Hall LA, Beissinger SR (2014) A practical toolbox for design and analysis of landscape genetics studies. *Landscape Ecology*, **29**, 1487–1504.

Helfer V, Broquet T, Fumagalli L (2012) Sex-specific estimates of dispersal show female philopatry and male dispersal in a promiscuous amphibian, the alpine salamander (*Salamandra atra*). *Molecular Ecology*, **21**, 4706–4720.

Hewitt GM (2001) Speciation, hybrid zones and phylogeography - Or seeing genes in space and time. *Molecular Ecology*, **10**, 537–549.

Hewitt GM (2004) Genetic consequences of climatic oscillations in the Quaternary. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, **359**, 183–195; discussion 195.

Hewitt GM (2011a) Mediterranean Peninsulas: The Evolution of Hotspots. In: *Biodiversity hotspots*, pp. 123–147.

Hewitt GM (2011b) Quaternary phylogeography: The roots of hybrid zones. *Genetica*, **139**, 617–638.

Jones OR, Wang J (2010) COLONY: A program for parentage and sibship inference from multilocus genotype data. *Molecular Ecology Resources*, **10**, 551–555.

Jost L (2008) GST and its relatives do not measure differentiation. *Molecular Ecology*, **17**, 4015–4026.

Kalinowski ST (2005) HP-RARE 1.0: A computer program for performing rarefaction on measures of allelic richness. *Molecular Ecology Notes*, **5**, 187–189.

Kaliontzopoulou A, Pinho C, Harris DJ, Carretero MA (2011) When cryptic diversity blurs the picture: A cautionary tale from Iberian and North African Podarcis wall lizards. *Biological Journal of the Linnean Society*, **103**, 779–800.

Librado P, Rozas J (2009) DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, **25**, 1451–1452.

Liu C, Berry PM, Dawson TP, Person RG (2005) Selecting Thresholds of Occurrence in the Predictions of Species Distribution. *Ecography*, **28**, 385–393.

Liu C, White M, Newell G (2013) Selecting thresholds for the prediction of species occurrence with presence-only data. *Journal of Biogeography*, **40**, 778–789.

Manel S, Schwartz MK, Luikart G, Taberlet P (2003) Landscape genetics: Combining landscape ecology and population genetics. *Trends in Ecology and Evolution*, **18**, 189–197.

Martínez-Solano I, Gonçalves H a., Arntzen JW, García-París M (2004) Phylogenetic relationships and biogeography of midwife toads (Discoglossidae: *Alytes*). *Journal of Biogeography*, **31**, 603–618.

Martínez-Solano I, Teixeira J, Buckley D, García-París M (2006) Mitochondrial DNA phylogeography of Lissotriton boscai (Caudata, Salamandridae): Evidence for old, multiple refugia in an Iberian endemic. *Molecular Ecology*, **15**, 3375–3388.

Martins IS, Proença V, Pereira HM (2014) The unusual suspect: Land use is a key predictor of biodiversity patterns in the Iberian Peninsula. *Acta Oecologica*, **61**, 41–50.

Mayden RL (1997) A hierarchy of species concepts: the denouement in the saga of the

species problem. *Species. The units of biodiversity.*, 381–423.

McRae BH (2006) Isolation By Resistance. *Evolution*, **60**, 1551–1561.

Miraldo A, Hewitt GM, Paulo OS, Emerson BC (2011) Phylogeography and demographic history of Lacerta lepida in the Iberian Peninsula: multiple refugia, range expansions and secondary contact zones. *BMC evolutionary biology*, **11**, 170.

Oksanen J (2016) Vegan: an introduction to ordination. *Management*, **1**, 1–10.

Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004) MICRO-CHECKER: Software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes*, **4**, 535–538.

Otto-Bliesner BL, Marshall SJ, Overpeck JT, Miller GH, Hu A (2006) Simulating Arctic climate warmth and icefield retreat in the last interglaciation. *Science (New York, N.Y.)*, **311**, 1751–1753.

Peakall R, Smouse PE (2006) GENALEX 6: Genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes*, **6**, 288–295.

Pereira RJ, Martínez-Solano I, Buckley D (2016) Hybridization during altitudinal range shifts: Nuclear introgression leads to extensive cyto-nuclear discordance in the fire salamander. *Molecular Ecology*, 1551–1565.

Peterman W, Brocato ER, Semlitsch RD, Eggert LS (2016) Reducing bias in population and landscape genetic inferences: the effects of sampling related individuals and multiple life stages. *PeerJ*, **4**, e1813.

Petit RJ, Excoffier L (2009) Gene flow and species delimitation. *Trends in Ecology and Evolution*, **24**, 386–393.

Phillips SB, Aneja VP, Kang D, Arya SP (2006) Modelling and analysis of the atmospheric nitrogen deposition in North Carolina. *International Journal of Global Environmental Issues*, **6**, 231–252.

Ponce de León JL, León G, Rodríguez R *et al.* (2014) Phylogeography of Cuban Rivulus: Evidence for allopatric speciation and secondary dispersal across a marine barrier. *Molecular Phylogenetics and Evolution*, **79**, 404–414.

Queiroz K De (2007) Species concepts and species delimitation. *Systematic Botany*, **56**, 879–886.

de Queiroz K (2005) Ernst Mayr and the modern concept of species. *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 6600–6607.

Queller DC, Goodnight KF (1989) Estimating Relatedness Using Genetic Markers Author (s): David C . Queller and Keith F . Goodnight Reviewed work (s): Published by : Society for the Study of Evolution Stable URL : http://www.jstor.org/stable/2409206 . *Society*, **43**, 258–275.

Romero D, Olivero J, Real R (2013) Comparative assessment of different methods for

using land-cover variables for distribution modelling of *Salamandra salamandra longirotris*. *Environmental Conservation*, **40**, 48–59.

Rousset F (2008) GENEPOP'007: A complete re-implementation of the GENEPOP software for Windows and Linux. *Molecular Ecology Resources*, **8**, 103–106.

Santos X, Rato C, Carranza S, Carretero MA, Pleguezuelos JM (2012) Complex phylogeography in the Southern Smooth Snake (*Coronella girondica*) supported by mtDNA sequences. *Journal of Zoological Systematics and Evolutionary Research*, **50**, 210–219.

Schulte U, Küsters D, Steinfartz S (2007) A PIT tag based analysis of annual movement patterns of adult fire salamanders (*Salamandra salamandra*) in a Middle European habitat. *Amphibia-Reptilia*, **28**, 531–536.

Seehausen O, Butlin RK, Keller I *et al.* (2014) Genomics and the origin of species. *Nature reviews. Genetics*, **15**, 176–92.

Selkoe KA, Toonen RJ (2006) Microsatellites for ecologists: A practical guide to using and evaluating microsatellite markers. *Ecology Letters*, **9**, 615–629.

Sequeira F, Alexandrino J, Weiss S, Ferrand N (2008) Documenting the advantages and limitations of different classes of molecular markers in a well-established phylogeographic context: Lessons from the Iberian endemic Golden-striped salamander, *Chioglossa lusitanica* (Caudata: Salamandridae). *Biological Journal of the Linnean Society*, **95**, 371–387.

Sillero N, Brito JC, Skidmore AK, Toxopeus AG (2009) Biogeographical patterns derived from remote sensing variables : the amphibians and reptiles of the Iberian Peninsula. , **30**, 185–206.

Smouse PE, Long JC, Sokal RR (1986) Society of Systematic Biologists Multiple Regression and Correlation Extensions of the Mantel Test of Matrix Correspondence Extensions of the Multiple Regression and Correlation Mantel Test of Matrix Correspondence. , **35**, 627–632.

Soberón J (2007) Grinnellian and Eltonian niches and geographic distributions of species. *Ecology Letters*, **10**, 1115–1123.

Spear SF, Balkenhol N, Fortin MJ, McRae BH, Scribner K (2010) Use of resistance surfaces for landscape genetic studies: Considerations for parameterization and analysis. *Molecular Ecology*, **19**, 3576–3591.

Speybroeck J, Crochet P, Beukema W (2010) Species list of the European herpetofauna – a tentative update. *Zootaxa*, **2492**, 1–27.

Steinfartz S, Veith M, Tautz D (2000) Mitochondrial sequence analysis of Salamandra taxa suggests old splits of major lineages and postglacial recolonizations of Central Europe from distinct source populations of Salamandra salamandra. *Molecular*

Ecology, **9**, 397–410.

Toews DPL, Brelsford A (2012) The biogeography of mitochondrial and nuclear discordance in animals. *Molecular Ecology*, **21**, 3907–3930.

Velo-Antón G, Fahd S, Teixeira J, Fritz U, Pereira P (2015) Out of Africa: did *Emys orbicularis* occidentaliscross the Strait of Gibraltar twice? *Amphibia-Reptilia*, **36**, 133–140.

Velo-Antón G, García-París M, Galán P, Cordero Rivera A (2007) The evolution of viviparity in holocene islands: Ecological adaptation versus phylogenetic descent along the transition from aquatic to terrestrial environments. *Journal of Zoological Systematics and Evolutionary Research*, **45**, 345–352.

Velo-antón G, Godinho R, Harris DJ *et al.* (2012) Deep evolutionary lineages in a Western Mediterranean snake (*Vipera latastei/monticola* group) and high genetic structuring in Southern Iberian populations. *Molecular Phylogenetics and Evolution*, **65**, 965–973.

Vences M, Sanchez E, Hauswaldt JS *et al.* (2014) Nuclear and mitochondrial multilocus phylogeny and survey of alkaloid content in true salamanders of the genus *Salamandra* (Salamandridae). *Molecular Phylogenetics and Evolution*, **73**, 208–216.

Veríssimo J, Znari M, Stuckas H *et al.* (2016) Pleistocene diversification in Morocco and recent demographic expansion in the Mediterranean pond turtle *Mauremys leprosa*. *Biological Journal of the Linnean Society*.

Wan QH, Wu H, Fujihara T, Fang SG (2004) Which genetic marker for which conservation genetics issue? *Electrophoresis*, **25**, 2165–2176.

Wang IJ (2010) Recognizing the temporal distinctions between landscape genetics and phylogeography. *Molecular Ecology*, **19**, 2605–2608.

Wang IJ, Bradburd GS (2014) Isolation by environment. *Molecular Ecology*, **23**, 5649–5662.

Welch AJ, Bedoya-Reina OC, Carretero-Paulet L *et al.* (2014) Polar bears exhibit genome-wide signatures of bioenergetic adaptation to life in the arctic environment. *Genome Biology and Evolution*, **6**, 433–450.

Wiens JJ (2004) Speciation and Ecology Revisited: Phylogenetic Niche Conservatism and the Origin of Species. *Evolution Orr and Smith Schluter*, **58**, 193–197.

Wiens JJ, Graham CH (2005) Niche Conservatism: Integrating Evolution, Ecology, and Conservation Biology. *Annual Review of Ecology, Evolution, and Systematics*, **36**, 519–539.

Zhang D-X, Hewitt GM (2003) Nuclear DNA analyses in genetic studies of populations: practice, problems and prospects. *Molecular ecology*, **12**, 563–84.

Zhang P, Papenfuss TJ, Wake MH, Qu L, Wake DB (2008) Phylogeny and biogeography of the family Salamandridae (Amphibia: Caudata) inferred from complete mitochondrial genomes. *Molecular Phylogenetics and Evolution*, **49**, 586–597.

6. Supplementary material

Appendix 1

Normalized Difference Vegetation Index (NDVI), obtained from pre-calculated NDVI layers at ~250 x 250 m resolution acquired from the MOD13Q1 product of the MODIS Terra satellite, available at NASA's Land Processes Distributed Active Archive Center (LPDAAC, https://lpdaac/usgs.gov). NDVI layers were obtained for 15 days intervals spanning the last 15 years (2000-2015), to allow characterization of annual and seasonal variation in vegetation cover, as well as to minimize the effect of extremes of vegetation abundance. GIS layers describing temporal patterns of variation vegetation were created on ArcMap (ESRI).

Table S1 - Details of the 9 microsatellites used in this study. Information regarding multiplex arrangement, original published primers and fluorescently labelled oligonucleotides used as template for modified forward primers are displayed. The primer volume used to create a multiplex with a total volume of 100 µl (distilled H_2O plus primer and fluorescent label's volumes) is also represented (PVM). In this process, the forward and reverse primers were concentrated at 10 µM and 100 µM respectively.

Locus	Multiplex	Label*	Primer forward (5' – 3')	Primer reverse (5' – 3')	PVM (µl)
Sal29	Panel S2	6-FAM	CTCTTTGACTGAACCAGAACCCC	GCCTGTCGGCTCTGTGTAACC	8.0
Sal23	Panel S3	6-FAM	TCACTGTTTATCTTTGTTCTTTTAT	AATTATTTGTTTGAGTCGATTTTCT	9.2
SalE7	Panel S4	NED	TTTCAGCACCAAGATACCTCTTTTG	CTCCCTCCATATCAAGGTCACAGAC	0.8
SalE5	Panel S4	6-FAM	CCACATGATGCCTACGTATGTTGTG	CTCCTGTTTACGCTTCACCTGCTCC	0.6
SalE2	Panel S4	VIC	CACGACAAAATACAGAGAGTGGATA	ATATTTGAAATTGCCCATTTGGTA	3.0
SalE06	Panel S5	VIC	GGACTCATGGTCACCCAGAGGTTCT	ATGGATTGTGTCGAAATAAGGTATC	1.2
Sal3	Panel S5	6-FAM	CTCAGACAAGAAATCCTGCTTCTTC	ATAAATCTGTCCTGTTCCTAATCAG	1.2
SalE8	Panel S5	NED	GCAAAGTCCATGCTTTCCCTTTCTC	GACATACCAAAGACTCCAGAATGGG	0.8
SalE12	Panel S3	VIC	CTCAGGAACAGTGTGCCCCAAATAC	CTCATAATTTAGTCTACCCTCCCAC	0.8

Table S2 - Information from STRUCTURE analyses of subset 3.

K	Reps	Mean LnP(K)	Stdev LnP(K)	Ln'(K)	|Ln''(K)|	Delta K
1	10	-10897.3	0.177951	—	—	—
2	10	-9665.76	0.333999	1231.49	689.82	2065.333
3	10	-9124.09	0.354181	541.67	203.69	575.1008
4	10	-8786.11	28.36169	337.98	182.84	6.446724
5	10	-8630.97	7.062113	155.14	17.33	2.45394
6	10	-8493.16	7.399429	137.81	117.84	15.92555
7	10	-8473.19	436.24	19.97	217.1	0.497662
8	10	-8236.12	64.54695	237.07	135.28	2.095839
9	10	-8134.33	67.79032	101.79	104.05	1.53488
10	10	-8136.59	237.5789	-2.26	110.79	0.466329
11	10	-8028.06	96.69609	108.53	115.55	1.194981
12	10	-8035.08	320.9575	-7.02	45.21	0.14086
13	10	-7996.89	658.5206	38.19	160.26	0.243364
14	10	-8118.96	876.6569	-122.07	162.58	0.185455
15	10	-8078.45	630.8602	40.51	—	—

Table S3 - Information from STRUCTURE analyses of subset 4 plus scattered individuals.

K	Reps	Mean LnP(K)	Stdev LnP(K)	Ln'(K)	|Ln''(K)|	Delta K
1	10	-13393.83	0.245176	—	—	—
2	10	-12618.22	29.814307	775.61	104.42	3.502345
3	10	-11947.03	1.492984	671.19	325.29	217.87915
4	10	-11601.13	1.800031	345.9	185.1	102.83157
5	10	-11440.33	17.654653	160.8	13.33	0.755042
6	10	-11266.2	20.845037	174.13	56.7	2.720072
7	10	-11148.77	55.721491	117.43	68.74	1.233635
8	10	-11100.08	273.03281	48.69	127.51	0.467013
9	10	-10923.88	177.771	176.2	77.37	0.435223
10	10	-10825.05	105.75807	98.83	—	—

Table S4 - Information from STRUCTURE analyses of subset 5 plus scattered individuals.

K	Reps	Mean LnP(K)	Stdev LnP(K)	Ln'(K)	|Ln''(K)|	Delta K
1	10	-5146.75	0.585472	—	—	—
2	10	-4622.85	0.604152	523.9	254.54	421.3176
3	10	-4353.49	0.665749	269.36	75.74	113.7665
4	10	-4159.87	2.710084	193.62	57.41	21.18385
5	10	-4023.66	61.43582	136.21	20.71	0.3371
6	10	-3908.16	1.055357	115.5	42.56	40.3276
7	10	-3835.22	0.856089	72.94	—	—

Table S5 - Sampling information and genetic diversity values for each population, using subset 2.

Population	Locality	N	N_A	H_O	H_E	AR	P-AR	R
VAL	Évora, Valverde	21	8.44	0.72	0.73	5.50	0.16	0.27
POR	Évora; Portel	14	6.78	0.72	0.71	5.09	0.15	0.29
PAY	Huelva	14	8.67	0.82	0.78	6.05	0.17	0.14
RSF	Huelva; Rosal de la Frontera	12	6.56	0.74	0.71	5.28	0.43	0.26
SDC	Setúbal; Santiago do Cacém	13	8.67	0.75	0.79	6.20	0.56	0.14
CAR	Beja; S. Martinho da Amoreiras/Carniceiro	15	10.33	0.77	0.83	6.86	0.39	0.11
PRT	Tavira, Portela	15	9.22	0.78	0.77	6.16	0.24	0.23
SLG	Sanlucar de Guadiana	13	4.56	0.67	0.60	3.86	0.33	0.51
ALA	Huelva; Aldea del Cabezuelo, Alájar	11	8.00	0.70	0.80	6.50	0.56	0.07
ARA	Huelva; Aracena	21	10.33	0.72	0.79	6.17	0.19	0.14
CAS	Sevilla; Cazalla de la Sierra	18	6.33	0.75	0.69	4.79	0.00	0.31
PLA	Sevilla; Las Navas de la Concepcion	11	6.00	0.74	0.70	4.84	0.07	0.26
JAR	Córdova, Las Jaras	11	5.33	0.68	0.69	4.36	0.03	0.28
AND	Jaén; Andújar	19	6.33	0.64	0.68	4.56	0.05	0.30
ORC	Jaén; Fte. Peña Alta, Orcera	14	6.22	0.69	0.67	4.73	0.62	0.35
MSI	Cádiz, Medina Siotonia	11	4.56	0.63	0.64	3.89	0.12	0.39
PLAG	Cádiz; Picalho, Sª del Aljibe, Alcalá de los Gazules	18	6.33	0.65	0.70	4.80	0.15	0.32
LMO	Málaga, Laguna del Moral	10	6.33	0.68	0.73	5.38	0.03	0.22
AVI	Cádiz, Arroyo de Valdeinfierno	16	6.89	0.68	0.72	5.04	0.10	0.28
VLR	Cádiz, Villaluenga del Rosario - Llanos del Republicano	14	6.11	0.53	0.63	4.67	0.15	0.40
TOR	Málaga, Camino a la cima del Torrecilla	12	3.89	0.50	0.54	3.34	0.14	0.52
VVR	Málaga, Villanueva del Rosario	16	2.44	0.29	0.27	2.09	0.22	0.83

* Mitochondrial DNA lineages identified with colors. *Salamandra s. gallaica* in yellow, *S. s. crespoi* in blue, *S. s. morenica* in green and *S. s. longirostris* in red.

Table S6 - Microsatellite pairwise F_{ST} and Jost's D_{EST} values above and below the diagonal respectively. Using subset 2.

Fst/D_{EST}	VAL	POR	SDC	CAR	PRT	PAY	SLG	RSF	ALA	ARA	CAS	PLA	JAR	AND	ORC	MSI	PLAG	LMO	AVI	VLR	TOR	VVR
VAL		0.033	0.052	0.113	0.191	0.082	0.288	0.088	0.135	0.096	0.116	0.152	0.152	0.162	0.127	0.261	0.242	0.210	0.223	0.273	0.309	0.456
POR	0.101		0.058	0.113	0.197	0.096	0.293	0.099	0.150	0.116	0.132	0.164	0.173	0.171	0.157	0.273	0.243	0.216	0.225	0.281	0.309	0.478
SDC	0.200	0.221		0.077	0.137	0.072	0.238	0.107	0.093	0.081	0.097	0.107	0.119	0.127	0.126	0.207	0.182	0.160	0.171	0.236	0.277	0.438
CAR	0.530	0.524	0.470		0.063	0.088	0.153	0.113	0.074	0.079	0.147	0.116	0.129	0.127	0.147	0.180	0.155	0.126	0.156	0.189	0.240	0.404
PRT	0.795	0.806	0.680	0.334		0.122	0.113	0.178	0.111	0.112	0.184	0.139	0.135	0.154	0.189	0.198	0.173	0.144	0.177	0.230	0.287	0.438
PAY	0.313	0.368	0.355	0.513	0.567		0.194	0.085	0.073	0.061	0.065	0.078	0.083	0.097	0.092	0.191	0.188	0.155	0.176	0.225	0.248	0.440
SLG	0.918	0.901	0.837	0.545	0.321	0.623		0.268	0.198	0.198	0.270	0.217	0.229	0.232	0.282	0.289	0.276	0.244	0.270	0.318	0.339	0.528
RSF	0.287	0.317	0.441	0.546	0.731	0.329	0.799		0.125	0.093	0.129	0.148	0.160	0.158	0.141	0.259	0.227	0.197	0.215	0.275	0.314	0.488
ALA	0.600	0.674	0.545	0.505	0.580	0.394	0.678	0.567		0.019	0.134	0.087	0.099	0.118	0.128	0.184	0.170	0.145	0.159	0.205	0.234	0.442
ARA	0.374	0.457	0.401	0.445	0.515	0.282	0.659	0.366	0.096		0.095	0.084	0.077	0.101	0.098	0.214	0.186	0.169	0.183	0.220	0.246	0.408
CAS	0.358	0.400	0.351	0.639	0.684	0.220	0.762	0.394	0.529	0.340		0.043	0.064	0.062	0.103	0.256	0.237	0.217	0.225	0.272	0.321	0.472
PLA	0.519	0.554	0.435	0.544	0.533	0.294	0.592	0.494	0.368	0.324	0.117		0.035	0.022	0.108	0.237	0.208	0.193	0.199	0.264	0.304	0.479
JAR	0.498	0.564	0.465	0.585	0.491	0.301	0.610	0.523	0.407	0.280	0.169	0.098		0.034	0.123	0.264	0.237	0.209	0.213	0.276	0.328	0.477
AND	0.507	0.524	0.458	0.517	0.535	0.326	0.603	0.482	0.444	0.349	0.159	0.059	0.086		0.102	0.251	0.233	0.207	0.220	0.279	0.329	0.454
ORC	0.387	0.476	0.460	0.627	0.689	0.314	0.775	0.424	0.495	0.341	0.275	0.304	0.341	0.263		0.274	0.259	0.231	0.234	0.297	0.342	0.502
MSI	0.902	0.922	0.806	0.773	0.705	0.707	0.759	0.870	0.741	0.827	0.791	0.754	0.836	0.746	0.838		0.105	0.077	0.114	0.197	0.260	0.501
PLAG	0.889	0.869	0.757	0.710	0.655	0.760	0.794	0.811	0.735	0.754	0.788	0.709	0.808	0.743	0.863	0.277		0.019	0.092	0.117	0.245	0.402
LMO	0.853	0.863	0.779	0.689	0.623	0.716	0.756	0.788	0.758	0.799	0.794	0.748	0.785	0.723	0.836	0.225	0.058		0.040	0.102	0.230	0.429
AVI	0.850	0.839	0.756	0.774	0.723	0.754	0.828	0.805	0.735	0.793	0.781	0.715	0.748	0.729	0.794	0.324	0.282	0.134		0.136	0.223	0.391
VLR	0.902	0.902	0.884	0.763	0.802	0.808	0.842	0.893	0.771	0.793	0.814	0.824	0.845	0.818	0.893	0.497	0.302	0.288	0.377		0.295	0.449
TOR	0.894	0.849	0.882	0.824	0.879	0.744	0.757	0.869	0.716	0.759	0.853	0.811	0.876	0.863	0.903	0.580	0.607	0.605	0.561	0.665		0.554
VVR	0.956	0.928	0.908	0.893	0.902	0.916	0.886	0.918	0.869	0.891	0.910	0.873	0.836	0.834	0.930	0.846	0.699	0.749	0.674	0.694	0.865	

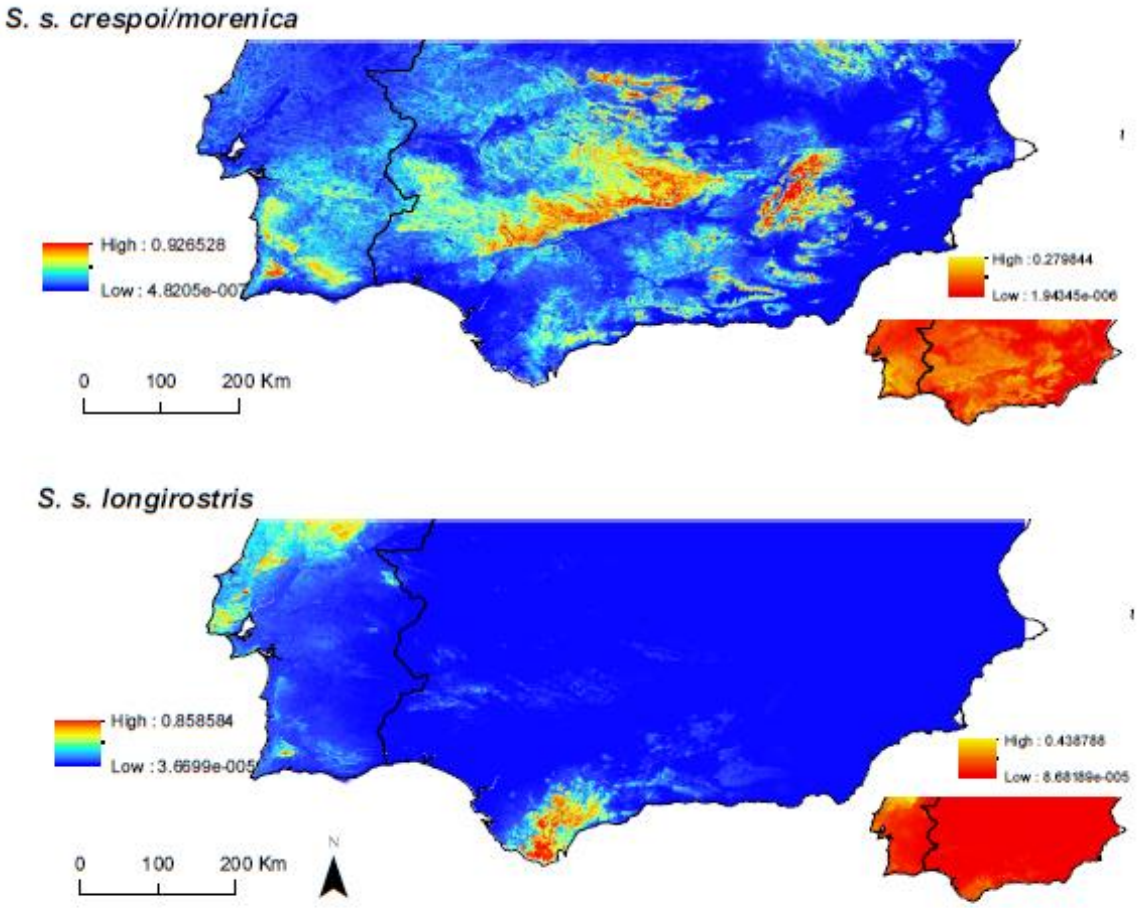


Figure S1. Projections of ecological niche models, showing suitability values, for S. s. longirostris and S. s. crespoi/morenica lineages, for the present, using 6 climatic variables (BIO3, Isothermality; BIO8, Mean Temperature of Wettest Quarter; BIO9, Mean Temperature of Driest Quarter; BIO11, Mean Temperature of Coldest Quarter; BIO12, Annual Precipitation; BIO18, Precipitation of Warmest Quarter), and 2 non-climatic variables (SLOPE and NDVI). Standard deviation of the models represented in the small inset.

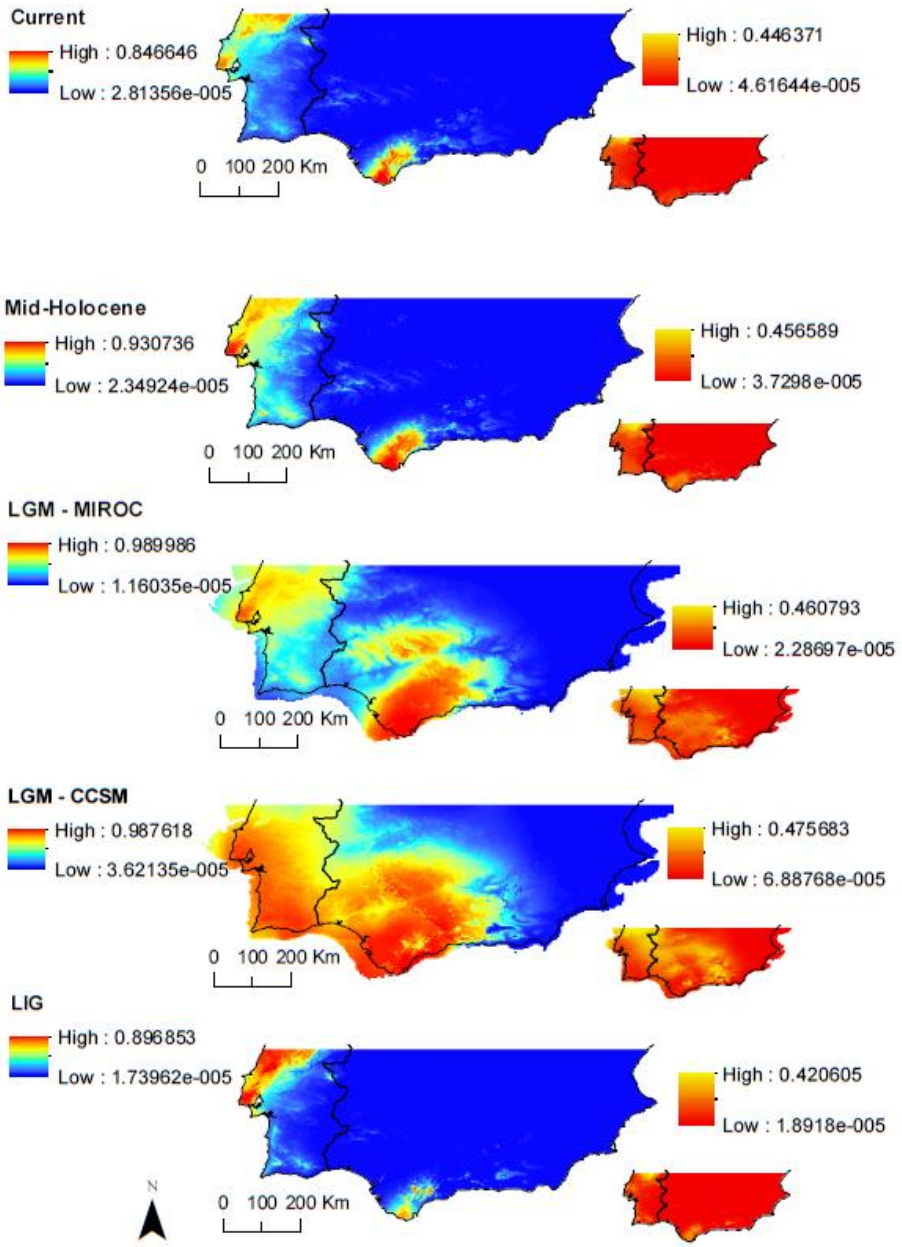


Figure S2. Projections of ecological niche models, showing suitability values, for *S. s. longirostris* lineage, for the present, Mid-Holocene, Last glacial maximum (CCSM and MIROC) and Last interglacial, using 6 climatic variables (BIO3, Isothermality; BIO8, Mean Temperature of Wettest Quarter; BIO9, Mean Temperature of Driest Quarter; BIO11, Mean Temperature of Coldest Quarter; BIO12, Annual Precipitation; BIO18, Precipitation of Warmest Quarter). Standard deviation of the models represented in the small insets.

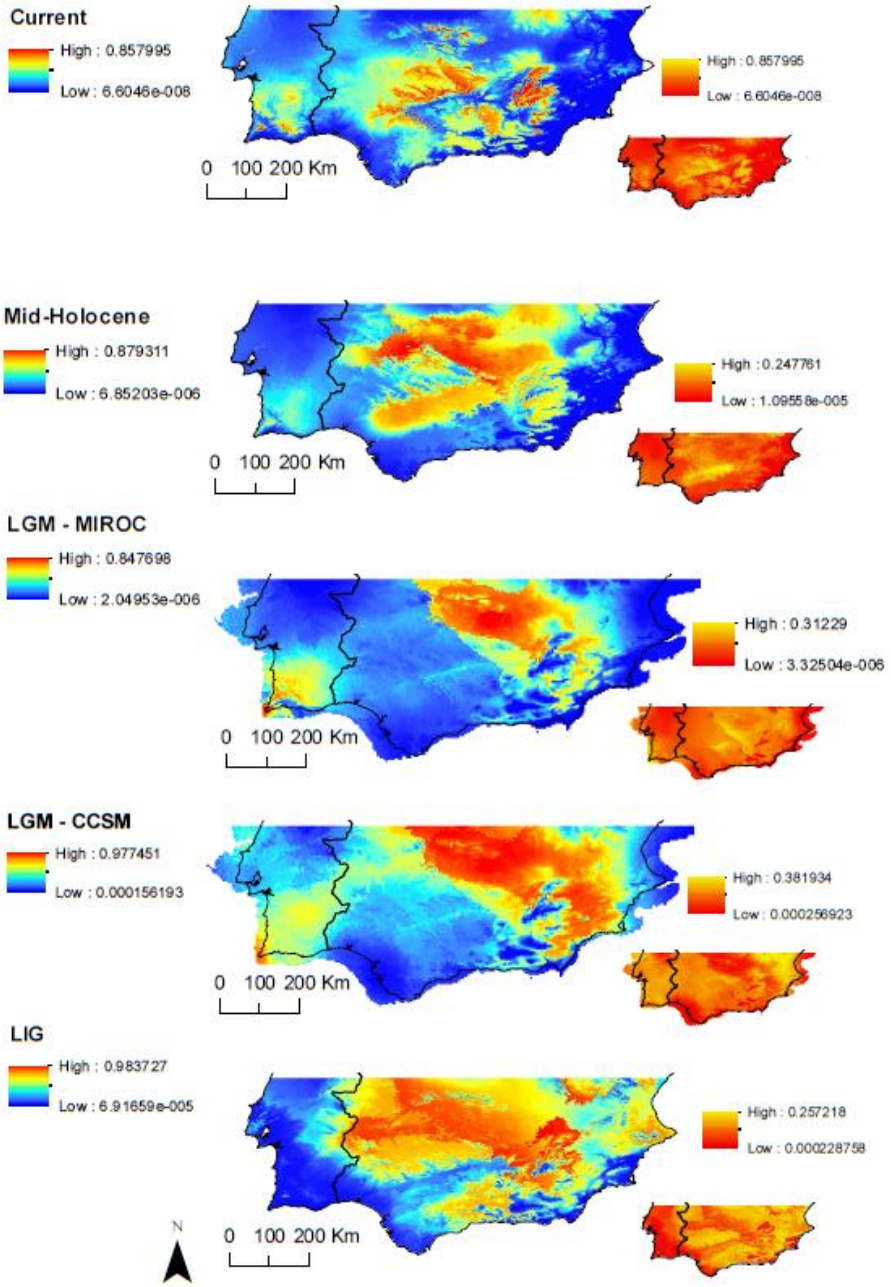


Figure S3. Projections of ecological niche models, showing suitability values, for *S. s. crespoi/morenica* lineage, for the present, Mid-Holocene, Last glacial maximum (CCSM and MIROC) and Last interglacial, using 6 climatic variables (BIO3, Isothermality; BIO8, Mean Temperature of Wettest Quarter; BIO9, Mean Temperature of Driest Quarter; BIO11, Mean Temperature of Coldest Quarter; BIO12, Annual Precipitation; BIO18, Precipitation of Warmest Quarter). Standard deviation of the models represented in the small inset.